


TRANSMITTAL OF UTILITY APPLICATION UNDER 37 C.F.R. § 1.53	Attorney Docket No.	24601-402F
	First named inventor	Hadlaczky
	Express mail label #	EL675147363US
	Date of mailing	November 28, 2000

Application Elements	Accompanying Application Papers
1. [X] Fee Transmittal Form	6. [X] Copy of assignments from parent application
2. [X] Specification containing <u>216</u> pages (including Sequence Listing, claims and Abstract).	7. [X] Small Entity Status is claimed
a. Title: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES	8. [] Preliminary Amendment
b. Number of claims: <u>33</u>	9. [X] Return Receipt Postcard
3. [X] <u>5</u> sheets of formal drawings with <u>5</u> Figs.	
4. [X] Copies of Declaration and Change of Power of Attorney by Assignee with Certificate Under 37 C.F.R. § 3.73(b) filed in parent application.	
5. [X] Sequence Listing	
[X] Paper copy (identical to computer copy)	
[X] Computer readable copy	
[] Verified statement	

SIGNATURE OF ATTORNEY/AGENT

HELLER EHRMAN WHITE & McAULIFFE LLP

Stephanie Seidman
Registration Number: 33,779

[X] Continuation of U.S. application Serial No. 08/835,682, filed April 10, 1997. This application is also a continuation-in-part of copending U.S. application Serial No. 08/695,191, filed August 7, 1996, now U.S. Patent No. 6,025,155, is also continuation-in-part of copending U.S. application Serial No. 08/682,080, filed July 15, 1996, now U.S. Patent No. 6,077,697 and is also a continuation-in-part of copending U.S. application Serial No. 08/629,822, filed April 10, 1996. The subject matter of each application is incorporated by reference.

CORRESPONDENCE ADDRESS

NAME	Stephanie Seidman Registration No. 33,779 Heller Ehrman White & McAuliffe LLP	
Address	4250 Executive Square, 7th Floor, La Jolla, CA 92037	
	Telephone: 858.450-8400	Facsimile: 858.587.5360

**FEE TRANSMITTAL
ACCOMPANYING UTILITY
APPLICATION UNDER
37 C.F.R. §1.53**

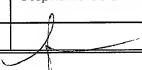
Attorney Docket No.	24601-402F
First named inventor	Hadlaczký
Express mail label #	EL675147363US
Date of mailing	November 28, 2000

FEE CALCULATION FOR CLAIMS AS AMENDED

a)	Basic Fee		\$	<u>710.00</u>
b)	Independent Claims $\frac{1}{3} - 3 = \frac{0}{3} \times \$ 80.00$		\$	<u>0.00</u>
c)	Total Claims $\frac{33}{33} - 20 = \frac{13}{33} \times \$ 18.00$		\$	<u>234.00</u>
d)	Fee for Multiple Dependent Claims - \$260.00		\$	<u>0.00</u>
	TOTAL FILING FEE		\$	<u>944.00</u>

- [X] Status as Small Entity is claimed,
reducing Fee by one-half to \$ 472.00
- [X] A check in the amount of \$472.00 to cover the fee for filing the application.
- [] Charge \$.00 to Deposit Account No. 50-1213.
- [X] The Commissioner is hereby authorized to charge any fees that may be required in this application under 37 C.F.R. §§ 1.16-1.17 during its entire pendency, or credit any overpayment, to Deposit Account No. 50-1213. If proper payment is not enclosed, such as a check in the wrong amount, unsigned, post-dated, otherwise improper or informal, or absent, the Commissioner is authorized to charge the unpaid amount to Deposit Account No. 50-1213 during the entire pendency of this application. This sheet is filed in duplicate.

CORRESPONDENCE ADDRESS

NAME	Stephanie Seidman Registration No. 33,779 Heller Ehrman White & McAuliffe				
Address	4250 Executive Square, 7th Floor, La Jolla, CA 92037				
	Telephone: 858.450.8400			Facsimile: 858.587.5860	
Submitted by:					
Typed or printed name	Stephanie Seidman			Reg. Number	33,779
Signature			Date	11/28/00	Deposit Account 50-1213

007210672 4420770

ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

RELATED APPLICATIONS

This application is a continuation of copending U.S. application

- 5 Serial No. 08/835,682, filed April 10, 1997, to GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also a continuation-in-part of copending U.S. application Serial No. 08/695,191, filed August 7, 1996, now U.S. Patent No. 6,025,155, to GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*. This application is also continuation-in-part of copending U.S. application Serial No. 08/682,080, filed July 15, 1996, now U.S. Patent No. 6,077,697, to GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*, and is also a continuation-in-part of copending U.S. application Serial No. 08/629,822, filed April 10, 1996 to GYULA HADLACZKY and ALADAR SZALAY, entitled *ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES*.
- 15
- 20

U.S. application Serial No. 08/835,682 is a continuation-in-part of U.S. application Serial No. 08/695,191. U.S. application Serial No. 08/695,191 is a continuation-in-part of U.S. application Serial No.

- 25 08/682,080 and also is a continuation-in-part of U.S. application Serial No. 08/629,822. U.S. application Serial No. 08/682,080 is a continuation-in-part of U.S. application Serial No. 08/629,822.

This application is related to U.S. application Serial No. 07/759,558, now U.S. Patent No. 5,288,625, is related to U.S.

- 30 application Serial No. 08/734,344, filed October 21, 1996, and is related to U.S. application Serial No. 08/375,271, filed 1/19/95, now U.S. Patent

- No. 5,712,134. U.S. application Serial No. 08/375,271 is a continuation of U.S. application Serial No. 08/080,097, filed 6/23/93 which is a continuation of U.S. application Serial No. 07/892,487, filed 6/3/92, which is a continuation of U.S. application Serial No. 07/521,073, filed 5/9/90.

The subject matter of each of the above-noted U.S. applications and patents is incorporated in its entirety by reference thereto.

FIELD OF THE INVENTION

- The present invention relates to methods for preparing cell lines that contain artificial chromosomes, methods for isolation of the artificial chromosomes, targeted insertion of heterologous DNA into the chromosomes, delivery of the chromosomes to selected cells and tissues and methods for isolation and large-scale production of the chromosomes. Also provided are cell lines for use in the methods, and cell lines and chromosomes produced by the methods. Further provided are cell-based methods for production of heterologous proteins, gene therapy methods and methods of generating transgenic animals, particularly non-human transgenic animals, that use artificial chromosomes.

BACKGROUND OF THE INVENTION

- Several viral vectors, non-viral, and physical delivery systems for gene therapy and recombinant expression of heterologous nucleic acids have been developed [see, e.g., Mitani *et al.* (1993) Trends Biotech. 11:162-166]. The presently available systems, however, have numerous limitations, particularly where persistent, stable, or controlled gene expression is required. These limitations include: (1) size limitations because there is a limit, generally on order of about ten kilobases [kB], at most, to the size of the DNA insert [gene] that can be accepted by viral vectors, whereas a number of mammalian genes of possible therapeutic importance are well above this limit, especially if all control elements are included; (2) the inability to specifically target integration so that random integration occurs which carries a risk of disrupting vital genes or cancer

suppressor genes; (3) the expression of randomly integrated therapeutic genes may be affected by the functional compartmentalization in the nucleus and are affected by chromatin-based position effects; (4) the copy number and consequently the expression of a given gene to be
5 integrated into the genome cannot be controlled. Thus, improvements in gene delivery and stable expression systems are needed [see, e.g., Mulligan (1993) Science 260:926-932].

In addition, safe and effective vectors and gene therapy methods should have numerous features that are not assured by the presently
10 available systems. For example, a safe vector should not contain DNA elements that can promote unwanted changes by recombination or mutation in the host genetic material, should not have the potential to initiate deleterious effects in cells, tissues, or organisms carrying the vector, and should not interfere with genomic functions. In addition, it
15 would be advantageous for the vector to be non-integrative, or designed for site-specific integration. Also, the copy number of therapeutic gene(s) carried by the vector should be controlled and stable, the vector should secure the independent and controlled function of the introduced gene(s); and the vector should accept large (up to Mb size) inserts and ensure the
20 functional stability of the insert.

The limitations of existing gene delivery technologies, however, argue for the development of alternative vector systems suitable for transferring large [up to Mb size or larger] genes and gene complexes together with regulatory elements that will provide a safe, controlled, and
25 persistent expression of the therapeutic genetic material.

At the present time, none of the available vectors fulfill all these requirements. Most of these characteristics, however, are possessed by chromosomes. Thus, an artificial chromosome would be an ideal vector for gene therapy, as well as for stable, high-level, controlled production of
30 gene products that require coordination of expression of numerous genes or that are encoded by large genes, and other uses. Artificial

- chromosomes for expression of heterologous genes in yeast are available, but construction of defined mammalian artificial chromosomes has not been achieved. Such construction has been hindered by the lack of an isolated, functional, mammalian centromere and uncertainty regarding the
- 5 requisites for its production and stable replication. Unlike in yeast, there are no selectable genes in close proximity to a mammalian centromere, and the presence of long runs of highly repetitive pericentric heterochromatic DNA makes the isolation of a mammalian centromere using presently available methods, such as chromosome walking, virtually
- 10 impossible. Other strategies are required for production of mammalian artificial chromosomes, and some have been developed. For example, U.S. Patent No. 5,288,625 provides a cell line that contains an artificial chromosome, a minichromosome, that is about 20 to 30 megabases. Methods provided for isolation of these chromosomes, however, provide
- 15 preparations of only about 10-20% purity. Thus, development of alternative artificial chromosomes and perfection of isolation and purification methods as well as development of more versatile chromosomes and further characterization of the minichromosomes is required to realize the potential of this technology.
- 20 Therefore, it is an object herein to provide mammalian artificial chromosomes and methods for introduction of foreign DNA into such chromosomes. It is also an object herein to provide methods of isolation and purification of the chromosomes. It is also an object herein to provide methods for introduction of the mammalian artificial chromosome
- 25 into selected cells, and to provide the resulting cells, as well as transgenic non-human animals, birds, fish and plants that contain the artificial chromosomes. It is also an object herein to provide methods for gene therapy and expression of gene products using artificial chromosomes. It is a further object herein to provide methods for constructing species-specific artificial chromosomes de novo. Another object herein is to provide methods to generate de novo mammalian artificial chromosomes.
- 30

SUMMARY OF THE INVENTION

Mammalian artificial chromosomes [MACs] are provided. Also provided are artificial chromosomes for other higher eukaryotic species, such as insects, birds, fowl and fish, produced using the MACS and methods provided herein. Methods for generating and isolating such chromosomes are provided. Methods using the MACs to construct artificial chromosomes from other species, such as insect, bird, fowl and fish species are also provided. The artificial chromosomes are fully functional stable chromosomes. Two types of artificial chromosomes are provided. One type, herein referred to as SATACs [satellite artificial chromosomes or satellite DNA based artificial chromosomes (the terms are used interchangeably herein)] are stable heterochromatic chromosomes, and the other type are minichromosomes based on amplification of euchromatin.

Artificial chromosomes provide an extra-genomic locus for targeted integration of megabase [Mb] pair size DNA fragments that contain single or multiple genes, including multiple copies of a single gene operatively linked to one promoter or each copy or several copies linked to separate promoters. Thus, methods using the MACs to introduce the genes into cells, tissues, and animals, as well as species such as birds, fowl, fish and plants, are also provided. The artificial chromosomes with integrated heterologous DNA may be used in methods of gene therapy, in methods of production of gene products, particularly products that require expression of multigenic biosynthetic pathways, and also are intended for delivery into the nuclei of germline cells, such as embryo-derived stem cells [ES cells], for production of transgenic (non-human) animals, birds, fowl and fish. Transgenic plants, including monocots and dicots, are also contemplated herein.

Mammalian artificial chromosomes provide extra-genomic specific integration sites for introduction of genes encoding proteins of interest and permit megabase size DNA integration so that, for example, genes

encoding an entire metabolic pathway or a very large gene, such as the cystic fibrosis [CF; ~250 kb] genomic DNA gene, several genes, such as multiple genes encoding a series of antigens for preparation of a multivalent vaccine, can be stably introduced into a cell. Vectors for

- 5 targeted introduction of such genes, including the tumor suppressor genes, such as p53, the cystic fibrosis transmembrane regulator cDNA [CFTR], and the genes for anti-HIV ribozymes, such as an anti-HIV gag ribozyme gene, into the artificial chromosomes are also provided.

- The chromosomes provided herein are generated by introducing
- 10 heterologous DNA that includes DNA encoding one or multiple selectable marker(s) into cells, preferably a stable cell line, growing the cells under selective conditions, and identifying from among the resulting clones those that include chromosomes with more than one centromere and/or fragments thereof. The amplification that produces the additional
- 15 centromere or centromeres occurs in cells that contain chromosomes in which the heterologous DNA has integrated near the centromere in the pericentric region of the chromosome. The selected clonal cells are then used to generate artificial chromosomes.

- Although non-targeted introduction of DNA, which results in some
- 20 frequency of integration into appropriate loci, targeted introduction is preferred. Hence, in preferred embodiments, the DNA with the selectable marker that is introduced into cells to initiate generation of artificial chromosomes includes sequences that target it to the an amplifiable region, such as the pericentric region, heterochromatin, and particularly
- 25 rDNA of the chromosome. For example, vectors, such as pTEMPUD and pHASPUd [provided herein], which include such DNA specific for mouse satellite DNA and human satellite DNA, respectively, are provided. The plasmid pHASPUd is a derivative of pTEMPUD that contains human satellite DNA sequences that specifically target human chromosomes.
- 30 Preferred targeting sequences include mammalian ribosomal RNA (rRNA) gene sequences (referred to herein as rDNA) which target the

heterologous DNA to integrate into the rDNA region of those chromosomes that contain rDNA. For example, vectors, such as pTERPUD, which include mouse rDNA, are provided. Upon integration into existing chromosomes in the cells, these vectors can induce the

- 5 amplification that results in generation of additional centromeres.

Artificial chromosomes are generated by culturing the cells with the multicentric, typically dicentric, chromosomes under conditions whereby the chromosome breaks to form a minichromosome and formerly dicentric chromosome. Among the MACs provided herein are the SATACs, which

- 10 are primarily made up of repeating units of short satellite DNA and are nearly fully heterochromatic, so that without insertion of heterologous or foreign DNA, the chromosomes preferably contain no genetic information or contain only non-protein-encoding gene sequences such as rDNA sequences. They can thus be used as "safe" vectors for delivery of DNA
15 to mammalian hosts because they do not contain any potentially harmful genes. The SATACs are generated, not from the minichromosome fragment as, for example, in U.S. Patent No. 5,288,625, but from the fragment of the formerly dicentric chromosome.

- In addition, methods for generating euchromatic minichromosomes
20 and the use thereof are also provided herein. Methods for generating one type of MAC, the minichromosome, previously described in U.S. Patent No. 5,288,625, and the use thereof for expression of heterologous DNA are provided. In a particular method provided herein for generating a
25 MAC, such as a minichromosome, heterologous DNA that includes mammalian rDNA and one or more selectable marker genes is introduced into cells which are then grown under selective conditions. Resulting cells that contain chromosomes with more than one centromere are selected and cultured under conditions whereby the chromosome breaks to form a minichromosome and a formerly multicentric (typically dicentric)
30 chromosome from which the minichromosome was released.

Cell lines containing the minichromosome and the use thereof for cell fusion are also provided. In one embodiment, a cell line containing the mammalian minichromosome is used as recipient cells for donor DNA encoding a selected gene or multiple genes. To facilitate integration of

5 the donor DNA into the minichromosome, the recipient cell line preferably contains the minichromosome but does not also contain the formerly dicentric chromosome. This may be accomplished by methods disclosed herein such as cell fusion and selection of cells that contain a minichromosome and no formerly dicentric chromosome. The donor DNA

10 is linked to a second selectable marker and is targeted to and integrated into the minichromosome. The resulting chromosome is transferred by cell fusion into an appropriate recipient cell line, such as a Chinese hamster cell line [CHO]. After large-scale production of the cells carrying the engineered chromosome, the chromosome is isolated. In particular,

15 metaphase chromosomes are obtained, such as by addition of colchicine, and they are purified from the cell lysate. These chromosomes are used for cloning, sequencing and for delivery of heterologous DNA into cells.

Also provided are SATACs of various sizes that are formed by repeated culturing under selective conditions and subcloning of cells that

20 contain chromosomes produced from the formerly dicentric chromosomes. The exemplified SATACs are based on repeating DNA units that are about 15 Mb [two ~7.5 Mb blocks]. The repeating DNA unit of SATACs formed from other species and other chromosomes may vary, but typically would be on the order of about 7 to about 20 Mb. The

25 repeating DNA units are referred to herein as megareplicons, which in the exemplified SATACs contain tandem blocks of satellite DNA flanked by non-satellite DNA, including heterologous DNA and non-satellite DNA. Amplification produces an array of chromosome segments [each called an amplicon] that contain two inverted megareplicons bordered by

30 heterologous ["foreign"] DNA. Repeated cell fusion, growth on selective medium and/or BrdU [5-bromodeoxyuridine] treatment or other treatment

with other genome destabilizing reagent or agent, such as ionizing radiation, including X-rays, and subcloning results in cell lines that carry stable heterochromatic or partially heterochromatic chromosomes, including a 150-200 Mb "sausage" chromosome, a 500-1000 Mb gigachromosome, a stable 250-400 Mb megachromosome and various smaller stable chromosomes derived therefrom. These chromosomes are based on these repeating units and can include heterologous DNA that is expressed.

Thus, methods for producing MACs of both types (i.e., SATACS and minichromosomes) are provided. These methods are applicable to the production of artificial chromosomes containing centromeres derived from any higher eukaryotic cell, including mammals, birds, fowl, fish, insects and plants.

The resulting chromosomes can be purified by methods provided herein to provide vectors for introduction of heterologous DNA into selected cells for production of the gene product(s) encoded by the heterologous DNA, for production of transgenic (non-human) animals, birds, fowl, fish and plants or for gene therapy.

In addition, methods and vectors for fragmenting the minichromosomes and SATACs are provided. Such methods and vectors can be used for in vivo generation of smaller stable artificial chromosomes. Vectors for chromosome fragmentation are used to produce an artificial chromosome that contains a megareplicon, a centromere and two telomeres and will be between about 7.5 Mb and about 60 Mb, preferably between about 10 Mb-15 Mb and 30-50 Mb. As exemplified herein, the preferred range is between about 7.5 Mb and 50 Mb. Such artificial chromosomes may also be produced by other methods.

Isolation of the 15 Mb [or 30 Mb amplicon containing two 15 Mb inverted repeats] or a 30 Mb or higher multimer, such as 60 Mb, thereof should provide a stable chromosomal vector that can be manipulated in

vitro. Methods for reducing the size of the MACs to generate smaller stable self-replicating artificial chromosomes are also provided.

- Also provided herein, are methods for producing mammalian artificial chromosomes, including those provided herein, in vitro, and the
- 5 resulting chromosomes. The methods involve *in vitro* assembly of the structural and functional elements to provide a stable artificial chromosome. Such elements include a centromere, two telomeres, at least one origin of replication and filler heterochromatin, e.g., satellite
- 10 DNA. A selectable marker for subsequent selection is also generally included. These specific DNA elements may be obtained from the artificial chromosomes provided herein such as those that have been generated by the introduction of heterologous DNA into cells and the subsequent amplification that leads to the artificial chromosome, particularly the SATACs. Centromere sequences for use in the *in vitro*
- 15 construction of artificial chromosomes may also be obtained by employing the centromere cloning methods provided herein. In preferred embodiments, the sequences providing the origin of replication, in particular, the megareplicator, are derived from rDNA. These sequences preferably include the rDNA origin of replication and amplification
- 20 promoting sequences.

Methods and vectors for targeting heterologous DNA into the artificial chromosomes are also provided as are methods and vectors for fragmenting the chromosomes to produce smaller but stable and self-replicating artificial chromosomes.

- 25 The chromosomes are introduced into cells to produce stable transformed cell lines or cells, depending upon the source of the cells. Introduction is effected by any suitable method including, but not limited to electroporation, direct uptake, such as by calcium phosphate precipitation, uptake of isolated chromosomes by lipofection, by microcell
- 30 fusion, by lipid-mediated carrier systems or other suitable method. The resulting cells can be used for production of proteins in the cells. The

chromosomes can be isolated and used for gene delivery. Methods for isolation of the chromosomes based on the DNA content of the chromosomes, which differs in MACs versus the authentic chromosomes, are provided. Also provided are methods that rely on content, particularly
5 density, and size of the MACs.

These artificial chromosomes can be used in gene therapy, gene product production systems, production of humanized genetically transformed animal organs, production of transgenic plants and animals (non-human), including mammals, birds, fowl, fish, invertebrates,
10 vertebrates, reptiles and insects, any organism or device that would employ chromosomal elements as information storage vehicles, and also for analysis and study of centromere function, for the production of artificial chromosome vectors that can be constructed in vitro, and for the preparation of species-specific artificial chromosomes. The artificial
15 chromosomes can be introduced into cells using microinjection, cell fusion, microcell fusion, electroporation, nuclear transfer, electrofusion, projectile bombardment, nuclear transfer, calcium phosphate precipitation, lipid-mediated transfer systems and other such methods. Cells particularly suited for use with the artificial chromosomes include, but are
20 not limited to plant cells, particularly tomato, arabidopsis, and others, insect cells, including silk worm cells, insect larvae, fish, reptiles, amphibians, arachnids, mammalian cells, avian cells, embryonic stem cells, haematopoietic stem cells, embryos and cells for use in methods of genetic therapy, such as lymphocytes that are used in methods of adop-
25 tive immunotherapy and nerve or neural cells. Thus methods of producing gene products and transgenic (non-human) animals and plants are provided. Also provided are the resulting transgenic animals and plants.

Exemplary cell lines that contain these chromosomes are also provided.

30 Methods for preparing artificial chromosomes for particular species and for cloning centromeres are also provided. For example, two

exemplary methods provided for generating artificial chromosomes for use in different species are as follows. First, the methods herein may be applied to different species. Second, means for generating species-specific artificial chromosomes and for cloning centromeres are provided.

- 5 In particular, a method for cloning a centromere from an animal or plant is provided by preparing a library of DNA fragments that contain the genome of the plant or animal and introducing each of the fragments into a mammalian satellite artificial chromosome [SATAC] that contains a centromere from a species, generally a mammal, different from the
10 selected plant or animal, generally a non-mammal, and a selectable marker. The selected plant or animal is one in which the mammalian species centromere does not function. Each of the SATACs is introduced into the cells, which are grown under selective conditions, and cells with SATACs are identified. Such SATACS should contain a centromere
15 encoded by the DNA from the library or should contain the necessary elements for stable replication in the selected species.

Also provided are libraries in which the relatively large fragments of DNA are contained on artificial chromosomes.

- Transgenic (non-human) animals, invertebrates and vertebrates,
20 plants and insects, fish, reptiles, amphibians, arachnids, birds, fowl, and mammals are also provided. Of particular interest are transgenic (non-human) animals and plants that express genes that confer resistance or reduce susceptibility to disease. For example, the transgene may encode a protein that is toxic to a pathogen, such as a virus, bacterium or pest,
25 but that is not toxic to the transgenic host. Furthermore, since multiple genes can be introduced on a MAC, a series of genes encoding an antigen can be introduced, which upon expression will serve to immunize [in a manner similar to a multivalent vaccine] the host animal against the diseases for which exposure to the antigens provide immunity or some
30 protection.

Also of interest are transgenic (non-human) animals that serve as models of certain diseases and disorders for use in studying the disease and developing therapeutic treatments and cures thereof. Such animal models of disease express genes [typically carrying a disease-associated mutation], which are introduced into the animal on a MAC and which induce the disease or disorder in the animal. Similarly, MACs carrying genes encoding antisense RNA may be introduced into animal cells to generate conditional "knock-out" transgenic (non-human) animals. In such animals, expression of the antisense RNA results in decreased or complete elimination of the products of genes corresponding to the antisense RNA. Of further interest are transgenic mammals that harbor MAC-carried genes encoding therapeutic proteins that are expressed in the animal's milk. Transgenic (non-human) animals for use in xenotransplantation, which express MAC-carried genes that serve to humanize the animal's organs, are also of interest. Genes that might be used in humanizing animal organs include those encoding human surface antigens.

Methods for cloning centromeres, such as mammalian centromeres, are also provided. In particular, in one embodiment, a library composed of fragments of SATACs are cloned into YACs [yeast artificial chromosomes] that include a detectable marker, such as DNA encoding tyrosinase, and then introduced into mammalian cells, such as albino mouse embryos. Mice produced from embryos containing such YACs that include a centromere that functions in mammals will express the detectable marker. Thus, if mice are produced from albino mouse embryos into which a functional mammalian centromere was introduced, the mice will be pigmented or have regions of pigmentation.

A method for producing repeated tandem arrays of DNA is provided. This method, exemplified herein using telomeric DNA, is applicable to any repeat sequence, and in particular, low complexity repeats. The method provided herein for synthesis of arrays of tandem

5

10

15

20

25

needed to achieve the desired repeat sequence size. In each extension cycle, the repeat sequence size doubles, i.e., if m is the number of extension cycles, the size of the repeat sequence will be $k \times 2^m$ nucleotides.

5 DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic drawing depicting formation of the MMCneo [the minichromosome] chromosome. A-G represents the successive events consistent with observed data that would lead to the formation and stabilization of the minichromosome.

- 10 Figure 2 shows a schematic summary of the manner in which the observed new chromosomes would form, and the relationships among the different *de novo* formed chromosomes. In particular, this figure shows a schematic drawing of the *de novo* chromosome formation initiated in the centromeric region of mouse chromosome 7. (A) A single E-type
- 15 amplification in the centromeric region of chromosome 7 generates a neo-centromere linked to the integrated "foreign" DNA, and forms a dicentric chromosome. Multiple E-type amplification forms the λ neo-chromosome, which separates from the remainder of mouse chromosome 7 through a specific breakage between the centromeres of the dicentric chromosome
- 20 and which was stabilized in a mouse-hamster hybrid cell line; (B) Specific breakage between the centromeres of a dicentric chromosome 7 generates a chromosome fragment with the neo-centromere, and a chromosome 7 with traces of heterologous DNA at the end; (C) Inverted duplication of the fragment bearing the neo-centromere results in the
- 25 formation of a stable neo-minichromosome; (D) Integration of exogenous DNA into the heterologous DNA region of the formerly dicentric chromosome 7 initiates H-type amplification, and the formation of a heterochromatic arm. By capturing a euchromatic terminal segment, this new chromosome arm is stabilized in the form of the "sausage"
- 30 chromosome; (E) BrdU [5-bromodeoxyuridine] treatment and/or drug selection induce further H-type amplification, which results in the

formation of an unstable gigachromosome: (F) Repeated BrdU treatments and/or drug selection induce further H-type amplification including a centromere duplication, which leads to the formation of another heterochromatic chromosome arm. It is split off from the chromosome 7
5 by chromosome breakage, and by acquiring a terminal segment, the stable megachromosome is formed.

Figure 3 is a schematic diagram of the replicon structure and a scheme by which a megachromosome could be produced.

Figure 4 sets forth the relationships among some of the exemplary
10 cell lines described herein.

Figure 5 is a diagram of the plasmid pTEMPUD.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Definitions

Unless defined otherwise, all technical and scientific terms used
15 herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications referred to herein are incorporated by reference.

As used herein, a mammalian artificial chromosome [MAC] is a piece of DNA that can stably replicate and segregate alongside
20 endogenous chromosomes. It has the capacity to accommodate and express heterologous genes inserted therein. It is referred to as a mammalian artificial chromosome because it includes an active mammalian centromere(s). Plant artificial chromosomes, insect artificial chromosomes and avian artificial chromosomes refer to chromosomes
25 that include plant and insect centromeres, respectively. A human artificial chromosome [HAC] refers to chromosomes that include human centromeres, BUGACs refer to insect artificial chromosomes, and AVACs refer to avian artificial chromosomes. Among the MACs provided herein

are SATACs, minichromosomes, and in vitro synthesized artificial chromosomes. Methods for construction of each type are provided herein.

- As used herein, in vitro synthesized artificial chromosomes are
- 5 artificial chromosomes that is produced by joining the essential components (at least the centromere, and origins of replication) in vitro.

As used herein, endogenous chromosomes refer to genomic chromosomes as found in the cell prior to generation or introduction of a MAC.

- As used herein, stable maintenance of chromosomes occurs when
- 10 at least about 85%, preferably 90%, more preferably 95%, of the cells retain the chromosome. Stability is measured in the presence of a selective agent. Preferably these chromosomes are also maintained in the absence of a selective agent. Stable chromosomes also retain their structure during cell culturing, suffering neither intrachromosomal nor
- 15 interchromosomal rearrangements.

As used herein, growth under selective conditions means growth of a cell under conditions that require expression of a selectable marker for survival.

- As used herein, an agent that destabilizes a chromosome is any
- 20 agent known by those of skill in the art to enhance amplification events, mutations. Such agents, which include BrdU, are well known to those of skill in the art.

- As used herein, *de novo* with reference to a centromere, refers to generation of an excess centromere as a result of incorporation of a
- 25 heterologous DNA fragment using the methods herein.

- As used herein, euchromatin and heterochromatin have their recognized meanings, euchromatin refers to chromatin that stains diffusely and that typically contains genes, and heterochromatin refers to chromatin that remains unusually condensed and that has been thought to
- 30 be transcriptionally inactive. Highly repetitive DNA sequences [satellite DNA], at least with respect to mammalian cells, are usually located in

regions of the heterochromatin surrounding the centromere [pericentric heterochromatin]. Constitutive heterochromatin refers to heterochromatin that contains the highly repetitive DNA which is constitutively condensed and genetically inactive.

- 5 As used herein, BrdU refers to 5-bromodeoxyuridine, which during replication is inserted in place of thymidine. BrdU is used as a mutagen; it also inhibits condensation of metaphase chromosomes during cell division.

- 10 As used herein, a dicentric chromosome is a chromosome that contains two centromeres. A multicentric chromosome contains more than two centromeres.

- 15 As used herein, a formerly dicentric chromosome is a chromosome that is produced when a dicentric chromosome fragments and acquires new telomeres so that two chromosomes, each having one of the centromeres, are produced. Each of the fragments are replicable chromosomes. If one of the chromosomes undergoes amplification of euchromatic DNA to produce a fully functional chromosome that contains the newly introduced heterologous DNA and primarily [at least more than 50%] euchromatin, it is a minichromosome. The remaining chromosome
- 20 is a formerly dicentric chromosome. If one of the chromosomes undergoes amplification, whereby heterochromatin [satellite DNA] is amplified and a euchromatic portion [or arm] remains, it is referred to as a sausage chromosome. A chromosome that is substantially all heterochromatin, except for portions of heterologous DNA, is called a
- 25 SATAC. Such chromosomes [SATACs] can be produced from sausage chromosomes by culturing the cell containing the sausage chromosome under conditions, such as BrdU treatment and/or growth under selective conditions, that destabilize the chromosome so that a satellite artificial chromosomes [SATAC] is produced. For purposes herein, it is
- 30 understood that SATACs may not necessarily be produced in multiple steps, but may appear after the initial introduction of the heterologous

DNA and growth under selective conditions, or they may appear after several cycles of growth under selective conditions and BrdU treatment.

As used herein, a SATAC refers to a chromosome that is substantially all heterochromatin, except for portions of heterologous

- 5 DNA. Typically, SATACs are satellite DNA based artificial chromosomes, but the term encompasses any chromosome made by the methods herein that contains more heterochromatin than euchromatin.

As used herein, amplifiable, when used in reference to a chromosome, particularly the method of generating SATACs provided

- 10 herein, refers to a region of a chromosome that is prone to amplification. Amplification typically occurs during replication and other cellular events involving recombination. Such regions are typically regions of the chromosome that include tandem repeats, such as satellite DNA, rDNA and other such sequences.

- 15 As used herein, amplification, with reference to DNA, is a process in which segments of DNA are duplicated to yield two or multiple copies of identical or nearly identical DNA segments that are typically joined as substantially tandem or successive repeats or inverted repeats.

- As used herein an amplicon is a repeated DNA amplification unit
- 20 that contains a set of inverted repeats of the megareplicon. A megareplicon represents a higher order replication unit. For example, with reference to the SATACs, the megareplicon contains a set of tandem DNA blocks each containing satellite DNA flanked by non-satellite DNA. Contained within the megareplicon is a primary replication site, referred to
- 25 as the megareplicator, which may be involved in organizing and facilitating replication of the pericentric heterochromatin and possibly the centromeres. Within the megareplicon there may be smaller [e.g., 50-300 kb in some mammalian cells] secondary replicons. In the exemplified SATACS, the megareplicon is defined by two tandem ~7.5 Mb DNA
- 30 blocks [see, e.g., Fig. 3]. Within each artificial chromosome [AC] or among a population thereof, each amplicon has the same gross structure

but may contain sequence variations. Such variations will arise as a result of movement of mobile genetic elements, deletions or insertions or mutations that arise, particularly in culture. Such variation does not affect the use of the ACs or their overall structure as described herein.

- 5 As used herein, ribosomal RNA [rRNA] is the specialized RNA that forms part of the structure of a ribosome and participates in the synthesis of proteins. Ribosomal RNA is produced by transcription of genes which, in eukaryotic cells, are present in multiple copies. In human cells, the approximately 250 copies of rRNA genes per haploid genome are spread
- 10 out in clusters on at least five different chromosomes (chromosomes 13, 14, 15, 21 and 22). In mouse cells, the presence of ribosomal DNA [rDNA] has been verified on at least 11 pairs out of 20 mouse chromosomes [chromosomes 5, 6, 9, 11, 12, 15, 16, 17, 18, 19 and X][see e.g., Rowe et al. (1996) Mamm. Genome 7:886-889 and Johnson
- 15 et al. (1993) Mamm. Genome 4:49-52]. In eukaryotic cells, the multiple copies of the highly conserved rRNA genes are located in a tandemly arranged series of rDNA units, which are generally about 40-45 kb in length and contain a transcribed region and a nontranscribed region known as spacer (i.e., intergenic spacer) DNA which can vary in length
- 20 and sequence. In the human and mouse, these tandem arrays of rDNA units are located adjacent to the pericentric satellite DNA sequences (heterochromatin). The regions of these chromosomes in which the rDNA is located are referred to as nucleolar organizing regions (NOR) which loop into the nucleolus, the site of ribosome production within the cell nucleus.

- 25 As used herein, the minichromosome refers to a chromosome derived from a multicentric, typically dicentric, chromosome [see, e.g., FIG. 1] that contains more euchromatic than heterochromatic DNA.

- As used herein, a megachromosome refers to a chromosome that, except for introduced heterologous DNA, is substantially composed of
- 30 heterochromatin. Megachromosomes are made of an array of repeated amplicons that contain two inverted megareplicons bordered by

introduced heterologous DNA [see, e.g., Figure 3 for a schematic drawing of a megachromosome]. For purposes herein, a megachromosome is about 50 to 400 Mb, generally about 250-400 Mb. Shorter variants are also referred to as truncated megachromosomes [about 90 to 120 or 150 Mb], dwarf megachromosomes [\sim 150-200 Mb] and cell lines, and a micro-megachromosome [\sim 50-90 Mb, typically 50-60 Mb]. For purposes herein, the term megachromosome refers to the overall repeated structure based on an array of repeated chromosomal segments [amplicons] that contain two inverted megareplicons bordered by any inserted heterologous DNA. The size will be specified.

As used herein, genetic therapy involves the transfer or insertion of heterologous DNA into certain cells, target cells, to produce specific gene products that are involved in correcting or modulating disease. The DNA is introduced into the selected target cells in a manner such that the heterologous DNA is expressed and a product encoded thereby is produced. Alternatively, the heterologous DNA may in some manner mediate expression of DNA that encodes the therapeutic product. It may encode a product, such as a peptide or RNA, that in some manner mediates, directly or indirectly, expression of a therapeutic product. Genetic therapy may also be used to introduce therapeutic compounds, such as TNF, that are not normally produced in the host or that are not produced in therapeutically effective amounts or at a therapeutically useful time. Expression of the heterologous DNA by the target cells within an organism afflicted with the disease thereby enables modulation of the disease. The heterologous DNA encoding the therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof.

As used herein, heterologous or foreign DNA and RNA are used interchangeably and refer to DNA or RNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations in the genome that differ from that in which it occurs in

nature. It is DNA or RNA that is not endogenous to the cell and has been exogenously introduced into the cell. Examples of heterologous DNA include, but are not limited to, DNA that encodes a gene product or gene product(s) of interest, introduced for purposes of gene therapy or for

5 production of an encoded protein. Other examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers drug resistance, DNA that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and DNA that encodes other types of proteins,

10 such as antibodies. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

As used herein, a therapeutically effective product is a product that is encoded by heterologous DNA that, upon introduction of the DNA into

15 a host, a product is expressed that effectively ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures said disease.

As used herein, transgenic plants refer to plants in which heterologous or foreign DNA is expressed or in which the expression of a

20 gene naturally present in the plant has been altered.

As used herein, operative linkage of heterologous DNA to regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional and translational stop sites, and other signal sequences refers to the relationship between such DNA and such

25 sequences of nucleotides. For example, operative linkage of heterologous DNA to a promoter refers to the physical relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes, binds to and transcribes the DNA in reading frame. Preferred promoters include

30 tissue specific promoters, such as mammary gland specific promoters,

viral promoters, such TK, CMV, adenovirus promoters, and other promoters known to those of skill in the art.

- As used herein, isolated, substantially pure DNA refers to DNA fragments purified according to standard techniques employed by those skilled in the art, such as that found in Maniatis et al. [(1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY].

- As used herein, expression refers to the process by which nucleic acid is transcribed into mRNA and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression may, if an appropriate eukaryotic host cell or organism is selected, include splicing of the mRNA.

- As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous DNA into cells for either expression of the heterologous DNA or for replication of the cloned heterologous DNA. Selection and use of such vectors and plasmids are well within the level of skill of the art.

- As used herein, transformation/transfection refers to the process by which DNA or RNA is introduced into cells. Transfection refers to the taking up of exogenous nucleic acid, e.g., an expression vector, by a host cell whether or not any coding sequences are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, by direct uptake using calcium phosphate [CaPO_4 ; see, e.g., Wigler et al. (1979) Proc. Natl. Acad. Sci. U.S.A., 76:1373-1376], polyethylene glycol [PEG]-mediated DNA uptake, electroporation, lipofection [see, e.g., Strauss (1996) Meth. Mol. Biol. 54:307-327], microcell fusion [see, EXAMPLES, see, also Lambert (1991) Proc. Natl. Acad. Sci. U.S.A. 88:5907-5911; U.S. Patent No. 5,396,767, Sawford et al. (1987) Somatic Cell Mol. Genet. 13:279-284; Dhar et al. (1984) Somatic Cell Mol. Genet. 10:547-559; and McNeill-Killary et al. (1995) Meth. Enzymol. 254:133-152], lipid-mediated carrier systems [see, e.g.,

Teifel et al. (1995) Biotechniques 19:79-80; Albrecht et al. (1996) Ann. Hematol. 72:73-79; Holmen et al. (1995) In Vitro Cell Dev. Biol. Anim. 31:347-351; REmy et al. (1994) Bioconjug. Chem. 5:647-654; Le Bolch et al. (1995) Tetrahedron Lett. 36:6681-6684; Loeffler et al. (1993)

- 5 Meth. Enzymol. 217:599-618] or other suitable method. Successful transfection is generally recognized by detection of the presence of the heterologous nucleic acid within the transfected cell, such as any indication of the operation of a vector within the host cell.
- Transformation means introducing DNA into an organism so that the DNA
- 10 is replicable, either as an extrachromosomal element or by chromosomal integration.

As used herein, injected refers to the microinjection [use of a small syringe] of DNA into a cell.

- As used herein, substantially homologous DNA refers to DNA that
- 15 includes a sequence of nucleotides that is sufficiently similar to another such sequence to form stable hybrids under specified conditions.

- It is well known to those of skill in this art that nucleic acid fragments with different sequences may, under the same conditions, hybridize detectably to the same "target" nucleic acid. Two nucleic acid
- 20 fragments hybridize detectably, under stringent conditions over a sufficiently long hybridization period, because one fragment contains a segment of at least about 14 nucleotides in a sequence which is complementary [or nearly complementary] to the sequence of at least one segment in the other nucleic acid fragment. If the time during which
- 25 hybridization is allowed to occur is held constant, at a value during which, under preselected stringency conditions, two nucleic acid fragments with exactly complementary base-pairing segments hybridize detectably to each other, departures from exact complementarity can be introduced into the base-pairing segments, and base-pairing will nonetheless occur to an
- 30 extent sufficient to make hybridization detectable. As the departure from complementarity between the base-pairing segments of two nucleic acids

becomes larger, and as conditions of the hybridization become more stringent, the probability decreases that the two segments will hybridize detectably to each other.

- Two single-stranded nucleic acid segments have "substantially the same sequence," within the meaning of the present specification, if (a) both form a base-paired duplex with the same segment, and (b) the melting temperatures of said two duplexes in a solution of 0.5 X SSPE differ by less than 10°C. If the segments being compared have the same number of bases, then to have "substantially the same sequence", they will typically differ in their sequences at fewer than 1 base in 10. Methods for determining melting temperatures of nucleic acid duplexes are well known [see, e.g., Meinkoth and Wahl (1984) Anal. Biochem. 138:267-284 and references cited therein].

- As used herein, a nucleic acid probe is a DNA or RNA fragment that includes a sufficient number of nucleotides to specifically hybridize to DNA or RNA that includes identical or closely related sequences of nucleotides. A probe may contain any number of nucleotides, from as few as about 10 and as many as hundreds of thousands of nucleotides. The conditions and protocols for such hybridization reactions are well known to those of skill in the art as are the effects of probe size, temperature, degree of mismatch, salt concentration and other parameters on the hybridization reaction. For example, the lower the temperature and higher the salt concentration at which the hybridization reaction is carried out, the greater the degree of mismatch that may be present in the hybrid molecules.

- To be used as a hybridization probe, the nucleic acid is generally rendered detectable by labelling it with a detectable moiety or label, such as ^{32}P , ^3H and ^{14}C , or by other means, including chemical labelling, such as by nick-translation in the presence of deoxyuridylate biotinylated at the 5'-position of the uracil moiety. The resulting probe includes the biotinylated uridylate in place of thymidylate residues and can be detected

- [via the biotin moieties] by any of a number of commercially available detection systems based on binding of streptavidin to the biotin. Such commercially available detection systems can be obtained, for example, from Enzo Biochemicals, Inc. [New York, NY]. Any other label known to
- 5 those of skill in the art, including non-radioactive labels, may be used as long as it renders the probes sufficiently detectable, which is a function of the sensitivity of the assay, the time available [for culturing cells, extracting DNA, and hybridization assays], the quantity of DNA or RNA available as a source of the probe, the particular label and the means used
- 10 to detect the label.

- Once sequences with a sufficiently high degree of homology to the probe are identified, they can readily be isolated by standard techniques, which are described, for example, by Maniatis et al. ((1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold
- 15 Spring Harbor, NY).

- As used herein, conditions under which DNA molecules form stable hybrids and are considered substantially homologous are such that DNA molecules with at least about 60% complementarity form stable hybrids. Such DNA fragments are herein considered to be "substantially
- 20 homologous". For example, DNA that encodes a particular protein is substantially homologous to another DNA fragment if the DNA forms stable hybrids such that the sequences of the fragments are at least about 60% complementary and if a protein encoded by the DNA retains its activity.

- 25 For purposes herein, the following stringency conditions are defined:

- 1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C
 - 2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C
 - 3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C
- 30 or any combination of salt and temperature and other reagents that result in selection of the same degree of mismatch or matching.

As used herein, immunoprotective refers to the ability of a vaccine or exposure to an antigen or immunity-inducing agent, to confer upon a host to whom the vaccine or antigen is administered or introduced, the ability to resist infection by a disease-causing pathogen or to have

5 reduced symptoms. The selected antigen is typically an antigen that is presented by the pathogen.

As used herein, all assays and procedures, such as hybridization reactions and antibody-antigen reactions, unless otherwise specified, are conducted under conditions recognized by those of skill in the art as

10 standard conditions.

A. Preparation of cell lines containing MACs

1. The megareplicon

The methods, cells and MACs provided herein are produced by virtue of the discovery of the existence of a higher-order replication unit

15 [megareplicon] of the centromeric region. This megareplicon is delimited by a primary replication initiation site [megareplicator], and appears to facilitate replication of the centromeric heterochromatin, and most likely, centromeres. Integration of heterologous DNA into the megareplicator region or in close proximity thereto, initiates a large-scale amplification of

20 megabase-size chromosomal segments, which leads to *de novo* chromosome formation in living cells.

DNA sequences that provide a preferred megareplicator are the rDNA units that give rise to ribosomal RNA (rRNA). In mammals, particularly mice and humans, these rDNA units contain specialized

25 elements, such as the origin of replication (or origin of bidirectional replication, *i.e.*, OBR, in mouse) and amplification promoting sequences (APS) and amplification control elements (ACE) (see, e.g., Gogel *et al.* (1996) Chromosoma 104:511-518; Coffman *et al.* (1993) Exp. Cell. Res. 209:123-132; Little *et al.* (1993) Mol. Cell. Biol. 13:6600-6613; Yoon *et al.* (1995) Mol. Cell. Biol. 15:2482-2489; Gonzalez and Sylvester (1995)

30

Genomics 27:320-328; Miesfeld and Arnheim (1982) Nuc. Acids Res. 10:3933-3949)); Maden et al. (1987) Biochem. J. 246:519-527).

- As described herein, without being bound by any theory, these specialized elements may facilitate replication and/or amplification of
- 5 megabase-size chromosomal segments in the *de novo* formation of chromosomes, such as those described herein, in cells. These specialized elements are typically located in the nontranscribed intergenic spacer region upstream of the transcribed region of rDNA. The intergenic spacer region may itself contain internally repeated sequences which can be
- 10 classified as tandemly repeated blocks and nontandem blocks (see e.g., Gonzalez and Sylvester (1995) Genomics 27:320-328). In mouse rDNA, an origin of bidirectional replication may be found within a 3-kb initiation zone centered approximately 1.6 kb upstream of the transcription start site (see, e.g., Gogel et al. (1996) Chromosoma 104:511-518). The
- 15 sequences of these specialized elements tend to have an altered chromatin structure, which may be detected, for example, by nuclease hypersensitivity or the presence of AT-rich regions that can give rise to bent DNA structures. An exemplary sequence encompassing an origin of replication is shown in SEQ ID NO. 16 and in GENBANK accession no.
- 20 X82564 at about positions 2430-5435. Exemplary sequences encompassing amplification-promoting sequences include nucleotides 690-1060 and 1105-1530 of SEQ ID NO. 16.

- In human rDNA, a primary replication initiation site may be found a few kilobase pairs upstream of the transcribed region and secondary
- 25 initiation sites may be found throughout the nontranscribed intergenic spacer region (see, e.g., Yoon et al. (1995) Mol. Cell. Biol. 15:2482-2489). A complete human rDNA repeat unit is presented in GENBANK as accession no. U13369 and is set forth in SEQ ID NO. 17 herein. Another exemplary sequence encompassing a replication initiation site
- 30 may be found within the sequence of nucleotides 35355-42486 in SEQ ID NO. 17 particularly within the sequence of nucleotides 37912-

42486 and more particularly within the sequence of nucleotides 37912-39288 of SEQ ID NO. 17 (see Coffman et al. (1993) Exp. Cell. Res. 209:123-132).

- Cell lines containing MACs can be prepared by transforming cells,
- 5 preferably a stable cell line, with a heterologous DNA fragment that encodes a selectable marker, culturing under selective conditions, and identifying cells that have a multicentric, typically dicentric, chromosome. These cells can then be manipulated as described herein to produce the minichromosomes and other MACs, particularly the heterochromatic
- 10 SATACs, as described herein.

- Development of a multicentric, particularly dicentric, chromosome typically is effected through integration of the heterologous DNA in the pericentric heterochromatin, preferably in the centromeric regions of chromosomes carrying rDNA sequences. Thus, the frequency of
- 15 incorporation can be increased by targeting to these regions, such as by including DNA, including, but not limited to, rDNA or satellite DNA, in the heterologous fragment that encodes the selectable marker. Among the preferred targeting sequences for directing the heterologous DNA to the pericentromeric heterochromatin are rDNA sequences that target
- 20 centromeric regions of chromosomes that carry rRNA genes. Such sequences include, but are not limited to, the DNA of SEQ ID NO. 16 and GENBANK accession no. X82564 and portions thereof, the DNA of SEQ ID NO. 17 and GENBANK accession no. U13369 and portions thereof, and the DNA of SEQ ID NOS. 18-24. A particular vector incorporating
- 25 from within SEQ ID NO. 16 for use in directing integration of heterologous DNA into chromosomal rDNA is pTERPUD (see Example 12). Satellite DNA sequences can also be used to direct the heterologous DNA to integrate into the pericentric heterochromatin. For example, vectors pTEMPUD and pHASPUd, which contain mouse and human satellite DNA,
- 30 respectively, are provided herein (see Example 12) as exemplary vectors

for introduction of heterologous DNA into cells for *de novo* artificial chromosome formation.

- The resulting cell lines can then be treated as the exemplified cells herein to produce cells in which the dicentric chromosome has
- 5 fragmented. The cells can then be used to introduce additional selective markers into the fragmented dicentric chromosome (i.e., formerly dicentric chromosome), whereby amplification of the pericentric heterochromatin will produce the heterochromatic chromosomes.

- The following discussion describes this process with reference to
- 10 the EC3/7 line and the resulting cells. The same procedures can be applied to any other cells, particularly cell lines to create SATACs and euchromatic minichromosomes.

2. Formation of *de novo* chromosomes

- De novo* centromere formation in a transformed mouse LMTK-fibro-
- 15 blast cell line [EC3/7] after cointegration of λ constructs [λ CM8 and λ gtWESneo] carrying human and bacterial DNA [Hadlaczy et al. (1991) Proc. Natl. Acad. Sci. U.S.A., 88:8106-8110 and U.S. application Serial No. 08/375,271] has been shown. The integration of the "heterologous" engineered human, bacterial and phage DNA, and the subsequent
- 20 amplification of mouse and heterologous DNA that led to the formation of a dicentric chromosome, occurred at the centromeric region of the short arm of a mouse chromosome. By G-banding, this chromosome was identified as mouse chromosome 7. Because of the presence of two functionally active centromeres on the same chromosome, regular
- 25 breakages occur between the centromeres. Such specific chromosome breakages gave rise to the appearance [in approximately 10% of the cells] of a chromosome fragment carrying the neo-centromere. From the EC3/7 cell line [see, U.S. Patent No. 5,288,625, deposited at the European Collection of Animal Cell Culture (hereinafter ECACC) under accession no.
- 30 90051001; see, also Hadlaczy et al. (1991) Proc. Natl. Acad. Sci. U.S.A., 88:8106-8110, and U.S. application Serial No. 08/375,271 and

the corresponding published European application EP 0 473 253, two
sublines [EC3/7C5 and EC3/7C6] were selected by repeated single-cell
cloning. In these cell lines, the neo-centromere was found exclusively on
a minichromosome [neo-minichromosome], while the formerly dicentric
5 chromosome carried traces of "heterologous" DNA.

It has now been discovered that integration of DNA encoding a
selectable marker in the heterochromatic region of the centromere led to
formation of the dicentric chromosome.

3. The neo-minichromosome

10 The chromosome breakage in the EC3/7 cells, which separates the
neo-centromere from the mouse chromosome, occurred in the G-band
positive "heterologous" DNA region. This is supported by the observation
of traces of λ and human DNA sequences at the broken end of the
formerly dicentric chromosome. Comparing the G-band pattern of the
15 chromosome fragment carrying the neo-centromere with that of the stable
neo-minichromosome, it is apparent that the neo-minichromosome is an
inverted duplicate of the chromosome fragment that bears the neo-
centromere. This is supported by the observation that although the neo-
minichromosome carries only one functional centromere, both ends of the
20 minichromosome are heterochromatic, and mouse satellite DNA
sequences were found in these heterochromatic regions by *in situ*
hybridization.

Mouse cells containing the minichromosome, which contains
multiple repeats of the heterologous DNA, which in the exemplified
25 embodiment is λ DNA and the neomycin-resistance gene, can be used as
recipient cells in cell transformation. Donor DNA, such as selected
heterologous DNA containing λ DNA linked to a second selectable marker,
such as the gene encoding hygromycin phosphotransferase which confers
hygromycin resistance [hyg], can be introduced into the mouse cells and
30 integrated into the minichromosomes by homologous recombination of λ
DNA in the donor DNA with that in the minichromosomes. Integration is

verified by *in situ* hybridization and Southern blot analyses. Transcription and translation of the heterologous DNA is confirmed by primer extension and immunoblot analyses.

- For example, DNA has been targeted into the neo-minichromosome
- 5 in EC3/7C5 cells using a λ DNA-containing construct [pNem1ruc] that also contains DNA encoding hygromycin resistance and the *Renilla* luciferase gene linked to a promoter, such as the cytomegalovirus [CMV] early promoter, and the bacterial neomycin resistance-encoding DNA. Integration of the donor DNA into the chromosome in selected cells
- 10 [designated PHN4] was confirmed by nucleic acid amplification [PCR] and *in situ* hybridization. Events that would produce a neo-minichromosome are depicted in Figure 1.

- The resulting engineered minichromosome that contains the heterologous DNA can then be transferred by cell fusion into a recipient
- 15 cell line, such as Chinese hamster ovary cells [CHO] and correct expression of the heterologous DNA can be verified. Following production of the cells, metaphase chromosomes are obtained, such as by addition of colchicine, and the chromosomes purified by addition of AT- and GC-specific dyes on a dual laser beam based cell sorter (see Example
- 20 10 B for a description of methods of isolating artificial chromomsomes). Preparative amounts of chromosomes [5×10^4 - 5×10^7 chromosomes/ml] at a purity of 95% or higher can be obtained. The resulting chromosomes are used for delivery to cells by methods such as microinjection and liposome-mediated transfer.

- 25 Thus, the neo-minichromosome is stably maintained in cells, replicates autonomously, and permits the persistent long-term expression of the *neo* gene under non-selective culture conditions. It also contains megabases of heterologous known DNA [λ DNA in the exemplified embodiments] that serves as target sites for homologous recombination
- 30 and integration of DNA of interest. The neo-minichromosome is, thus, a vector for genetic engineering of cells. It has been introduced into SCID

mice, and shown to replicate in the same manner as endogenous chromosomes.

- The methods herein provide means to induce the events that lead to formation of the neo-minichromosome by introducing heterologous
- 5 DNA with a selective marker [preferably a dominant selectable marker] into cells and culturing the cells under selective conditions. As a result, cells that contain a multicentric, e.g., dicentric chromosome, or fragments thereof, generated by amplification are produced. Cells with the dicentric chromosome can then be treated to destabilize the chromosomes with
 - 10 agents, such as BrdU and/or culturing under selective conditions, resulting in cells in which the dicentric chromosome has formed two chromosomes, a so-called minichromosome, and a formerly dicentric chromosome that has typically undergone amplification in the heterochromatin where the heterologous DNA has integrated to produce a SATAC or a sausage
 - 15 chromosome [discussed below]. These cells can be fused with other cells to separate the minichromosome from the formerly dicentric chromosome into different cells so that each type of MAC can be manipulated separately.

4. Preparation of SATACs

- 20 An exemplary protocol for preparation of SATACs is illustrated in Figure 2 [particularly D, E and F] and FIGURE 3 [see, also the EXAMPLES, particularly EXAMPLES 4-7].

- To prepare a SATAC, the starting materials are cells, preferably a stable cell line, such as a fibroblast cell line, and a DNA fragment that
- 25 includes DNA that encodes a selective marker. The DNA fragment is introduced into the cell by methods of DNA transfer, including but not limited to direct uptake using calcium phosphate, electroporation, and lipid-mediated transfer. To insure integration of the DNA fragment in the heterochromatin, it is preferable to start with DNA that will be targeted to
 - 30 the pericentric heterochromatic region of the chromosome, such as λ CM8 and vectors provided herein, such as pTEMPUD [Figure 5] and pHASPU

(see Example 12) that include satellite DNA, or specifically into rDNA in the centromeric regions of chromosomes containing rDNA sequences.

After introduction of the DNA, the cells are grown under selective conditions. The resulting cells are examined and any that have

- 5 multicentric, particularly dicentric, chromosomes [or heterochromatic chromosomes or sausage chromosomes or other such structure; see, Figure 2D, 2E and 2F] are selected.

In particular, if a cell with a dicentric chromosome is selected, it can be grown under selective conditions, or, preferably, additional DNA

- 10 encoding a second selectable marker is introduced, and the cells grown under conditions selective for the second marker. The resulting cells should include chromosomes that have structures similar to those depicted in Figures 2D, 2E, 2F. Cells with a structure, such as the sausage chromosome, Figure 2D, can be selected and fused with a
- 15 second cell line to eliminate other chromosomes that are not of interest. If desired, cells with other chromosomes can be selected and treated as described herein. If a cell with a sausage chromosome is selected, it can be treated with an agent, such as BrdU, that destabilizes the chromosome so that the heterochromatic arm forms a chromosome that is substantially
- 20 heterochromatic [i.e., a megachromosome, see, Figure 2F]. Structures such as the gigachromosome in which the heterochromatic arm has amplified but not broken off from the euchromatic arm, will also be observed. The megachromosome is a stable chromosome. Further manipulation, such as fusions and growth in selective conditions and/or
- 25 BrdU treatment or other such treatment, can lead to fragmentation of the megachromosome to form smaller chromosomes that have the amplicon as the basic repeating unit.

The megachromosome can be further fragmented in vivo using a chromosome fragmentation vector, such as pTEMPUD [see, Figure 5 and

- 30 EXAMPLE 12], pHASPUD or pTERPUD (see Example 12) to ultimately

produce a chromosome that comprises a smaller stable replicable unit, about 15 Mb-60 Mb, containing one to four megareplicons.

Thus, the stable chromosomes formed *de novo* that originate from the short arm of mouse chromosome 7 have been analyzed. This

- 5 chromosome region shows a capacity for amplification of large chromosome segments, and promotes *de novo* chromosome formation. Large-scale amplification at the same chromosome region leads to the formation of dicentric and multicentric chromosomes, a minichromosome, the 150-200 Mb size λ neo-chromosome, the "sausage" chromosome, the
- 10 500-1000 Mb gigachromosome, and the stable 250-400 Mb megachromosome.

A clear segmentation is observed along the arms of the megachromosome, and analyses show that the building units of this chromosome are amplicons of ~30 Mb composed of mouse major

- 15 satellite DNA with the integrated "foreign" DNA sequences at both ends. The ~30 Mb amplicons are composed of two ~15 Mb inverted doublets of ~7.5 Mb mouse major satellite DNA blocks, which are separated from each other by a narrow band of non-satellite sequences [see, e.g., Figure 3]. The wider non-satellite regions at the amplicon borders contain
- 20 integrated, exogenous [heterologous] DNA, while the narrow bands of non-satellite DNA sequences within the amplicons are integral parts of the pericentric heterochromatin of mouse chromosomes. These results indicate that the ~7.5 Mb blocks flanked by non-satellite DNA are the building units of the pericentric heterochromatin of mouse chromosomes,
- 25 and the ~15 Mb size pericentric regions of mouse chromosomes contain two ~7.5 Mb units.

Apart from the euchromatic terminal segments, the whole megachromosome is heterochromatic, and has structural homogeneity. Therefore, this large chromosome offers a unique possibility for obtaining

- 30 information about the amplification process, and for analyzing some basic

characteristics of the pericentric constitutive heterochromatin, as a vector for heterologous DNA, and as a target for further fragmentation.

As shown herein, this phenomenon is generalizable and can be observed with other chromosomes. Also, although these *de novo* formed chromosome segments and chromosomes appear different, there are similarities that indicate that a similar amplification mechanism plays a role in their formation: (i) in each case, the amplification is initiated in the centromeric region of the mouse chromosomes and large (Mb size) amplicons are formed; (ii) mouse major satellite DNA sequences are constant constituents of the amplicons, either by providing the bulk of the heterochromatic amplicons [H-type amplification], or by bordering the euchromatic amplicons [E-type amplification]; (iii) formation of inverted segments can be demonstrated in the λ neo-chromosome and megachromosome; (iv) chromosome arms and chromosomes formed by the amplification are stable and functional.

The presence of inverted chromosome segments seems to be a common phenomenon in the chromosomes formed *de novo* at the centromeric region of mouse chromosome 7. During the formation of the neo-minichromosome, the event leading to the stabilization of the distal segment of mouse chromosome 7 that bears the neo-centromere may have been the formation of its inverted duplicate. Amplicons of the megachromosome are inverted doublets of ~ 7.5 Mb mouse major satellite DNA blocks.

5. Cell lines

Cell lines that contain MACs, such as the minichromosome, the λ -neo chromosome, and the SATACs are provided herein or can be produced by the methods herein. Such cell lines provide a convenient source of these chromosomes and can be manipulated, such as by cell fusion or production of microcells for fusion with selected cell lines, to deliver the chromosome of interest into hybrid cell lines. Exemplary cell

lines are described herein and some have been deposited with the ECACC.

a. EC3/7C5 and EC3/7C6

- Cell lines EC3/7C5 and EC3/7C6 were produced by single cell
- 5 cloning of EC3/7. For exemplary purposes EC3/7C5 has been deposited with the ECACC. These cell lines contain a minichromosome and the formerly dicentric chromosome from EC3/7. The stable minichromosomes in cell lines EC3/7C5 and EC3/7C6 appear to be the same and they seem to be duplicated derivatives of the ~10-15 Mb "broken-off" fragment of
- 10 the dicentric chromosome. Their similar size in these independently generated cell lines might indicate that ~20-30 Mb is the minimal or close to the minimal physical size for a stable minichromosome.

b. TF1004G19

- Introduction of additional heterologous DNA, including DNA
- 15 encoding a second selectable marker, hygromycin phosphotransferase, i.e., the hygromycin-resistance gene, and also a detectable marker, β -galactosidase (i.e., encoded by the lacZ gene), into the EC3/7C5 cell line and growth under selective conditions produced cells designated TF1004G19. In particular, this cell line was produced from the EC3/7C5
- 20 cell line by cotransfection with plasmids pH132, which contains an anti-HIV ribozyme and hygromycin-resistance gene, pCH110 [encodes β -galactosidase] and λ phage [λ cl 875 Sam 7] DNA and selection with hygromycin B.

- Detailed analysis of the TF1004G19 cell line by *in situ* hybridization
- 25 with λ phage and plasmid DNA sequences revealed the formation of the sausage chromosome. The formerly dicentric chromosome of the EC3/7C5 cell line translocated to the end of another acrocentric chromosome. The heterologous DNA integrated into the pericentric heterochromatin of the formerly dicentric chromosome and is amplified
- 30 several times with megabases of mouse pericentric heterochromatic satellite DNA sequences [Fig. 2D] forming the "sausage" chromosome.

Subsequently the acrocentric mouse chromosome was substituted by a euchromatic telomere.

In situ hybridization with biotin-labeled subfragments of the hygromycin-resistance and β -galactosidase genes resulted in a

- 5 hybridization signal only in the heterochromatic arm of the sausage chromosome, indicating that in TF1004G19 transformant cells these genes are localized in the pericentric heterochromatin.

- A high level of gene expression, however, was detected. In general, heterochromatin has a silencing effect in *Drosophila*, yeast and
10 on the HSV-tk gene introduced into satellite DNA at the mouse centromere. Thus, it was of interest to study the TF1004G19 transformed cell line to confirm that genes located in the heterochromatin were indeed expressed, contrary to recognized dogma.

- For this purpose, subclones of TF1004G19, containing a different
15 sausage chromosome [see Figure 2D], were established by single cell cloning. Southern hybridization of DNA isolated from the subclones with subfragments of hygromycin phosphotransferase and lacZ genes showed a close correlation between the intensity of hybridization and the length of the sausage chromosome. This finding supports the conclusion that
20 these genes are localized in the heterochromatic arm of the sausage chromosome.

(1) TF1004G-19C5

- TF1004G-19C5 is a mouse LMTK⁻ fibroblast cell line containing neo-minichromosomes and stable "sausage" chromosomes. It is a
25 subclone of TF1004G19 and was generated by single-cell cloning of the TF1004G19 cell line. It has been deposited with the ECACC as an exemplary cell line and exemplary source of a sausage chromosome. Subsequent fusion of this cell line with CHO K20 cells and selection with hygromycin and G418 and HAT (hypoxanthine, aminopterin, and
30 thymidine medium; see Szybalski *et al.* (1962) Proc. Natl. Acad. Sci. 48:2026) resulted in hybrid cells (designated 19C5xHa4) that carry the

sausage chromosome and the neo-minichromosome. BrdU treatment of the hybrid cells, followed by single cell cloning and selection with G418 and/or hygromycin produced various cells that carry chromosomes of interest, including GB43 and G3D5.

5 (2) other subclones

- Cell lines GB43 and G3D5 were obtained by treating 19C5xHa4 cells with BrdU followed by growth in G418-containing selective medium and retreatment with BrdU. The two cell lines were isolated by single cell cloning of the selected cells. GB43 cells contain the neo-minichromosome only. G3D5, which has been deposited with the ECACC, carries the neo-minichromosome and the megachromosome. Single cell cloning of this cell line followed by growth of the subclones in G418- and hygromycin-containing medium yielded subclones such as the GHB42 cell line carrying the neo-minichromosome and the megachromosome. H1D3 is a mouse-hamster hybrid cell line carrying the megachromosome, but no neo-minichromosome, and was generated by treating 19C5xHa4 cells with BrdU followed by growth in hygromycin-containing selective medium and single cell subcloning of selected cells. Fusion of this cell line with the CD4⁺ HeLa cell line that also carries DNA encoding an additional selection gene, the neomycin-resistance gene, produced cells [designated H1xHE41 cells] that carry the megachromosome as well as a human chromosome that carries CD4neo. Further BrdU treatment and single cell cloning produced cell lines, such as 1B3, that include cells with a truncated megachromosome.

25 5. DNA constructs used to transform the cells

- Heterologous DNA can be introduced into the cells by transfection or other suitable method at any stage during preparation of the chromosomes [see, e.g., FIG. 4]. In general, incorporation of such DNA into the MACs is assured through site-directed integration, such as may be accomplished by inclusion of λ -DNA in the heterologous DNA (for the exemplified chromosomes), and also an additional selective marker gene.

For example, cells containing a MAC, such as the minichromosome or a SATAC, can be cotransfected with a plasmid carrying the desired heterologous DNA, such as DNA encoding an HIV ribozyme, the cystic fibrosis gene, and DNA encoding a second selectable marker, such as

5 hygromycin resistance. Selective pressure is then applied to the cells by exposing them to an agent that is harmful to cells that do not express the new selectable marker. In this manner, cells that include the heterologous DNA in the MAC are identified. Fusion with a second cell

10 line can provide a means to produce cell lines that contain one particular type of chromosomal structure or MAC.

Various vectors for this purpose are provided herein [see, Examples] and others can be readily constructed. The vectors preferably include DNA that is homologous to DNA contained within a MAC in order to target the DNA to the MAC for integration therein. The vectors also

15 include a selectable marker gene and the selected heterologous gene(s) of interest. Based on the disclosure herein and the knowledge of the skilled artisan, one of skill can construct such vectors.

Of particular interest herein is the vector pTEMPUD and derivatives thereof that can target DNA into the heterochromatic region of selected

20 chromosomes. These vectors can also serve as fragmentation vectors [see, e.g., Example 12].

Heterologous genes of interest include any gene that encodes a therapeutic product and DNA encoding gene products of interest. These genes and DNA include, but are not limited to: the cystic fibrosis gene

25 [CF], the cystic fibrosis transmembrane regulator (CFTR) gene [see, e.g., U.S. Patent No. 5,240,846; Rosenfeld et al. (1992) Cell 68:143-155; Hyde et al. (1993) Nature 362: 250-255; Kerem et al. (1989) Science 245:1073-1080; Riordan et al. (1989) Science 245:1066-1072; Rommens et al. (1989) Science 245:1059-1065; Osborne et al. (1991)

30 Am. J. Hum. Genetics 48:6089-6122; White et al. (1990) Nature 344:665-667; Dean et al. (1990) Cell 61:863-870; Erlich et al. (1991)

Science 252:1643; and U.S. Patent Nos. 5,453,357, 5,449,604, 5,434,086, and 5,240,846, which provides a retroviral vector encoding the normal CFTR gene].

B. Isolation of artificial chromosomes

- 5 The MACs provided herein can be isolated by any suitable method known to those of skill in the art. Also, methods are provided herein for effecting substantial purification, particularly of the SATACs. SATACs have been isolated by fluorescence-activated cell sorting [FACS]. This method takes advantage of the nucleotide base content of the SATACs,
- 10 which, by virtue of their high heterochromatic DNA content, will differ from any other chromosomes in a cell. In particular embodiment, metaphase chromosomes are isolated and stained with base-specific dyes, such as Hoechst 33258 and chromomycin A3. Fluorescence-activated cell sorting will separate the SATACs from the endogenous
- 15 chromosomes. A dual-laser cell sorter [FACS Vantage Becton Dickinson Immunocytometry Systems] in which two lasers were set to excite the dyes separately, allowed a bivariate analysis of the chromosomes by base-pair composition and size. Cells containing such SATACs can be similarly sorted.
- 20 Additional methods provided herein for isolation of artificial chromosomes from endogenous chromosomes include procedures that are particularly well suited for large-scale isolation of artificial chromosomes such as SATACs. In these methods, the size and density differences between SATACs and endogenous chromosomes are exploited to effect
- 25 separation of these two types of chromosomes. Such methods involve techniques such as swinging bucket centrifugation, zonal rotor centrifugation, and velocity sedimentation. Affinity-, particularly immunoaffinity-, based methods for separation of artificial from endogenous chromosomes are also provided herein. For example,
- 30 SATACs, which are predominantly heterochromatin, may be separated from endogenous chromosomes through immunoaffinity procedures

involving antibodies that specifically recognize heterochromatin, and/or the proteins associated therewith, when the endogenous chromosomes contain relatively little heterochromatin, such as in hamster cells.

C. *In vitro* construction of artificial chromosomes

- 5 Artificial chromosomes can be constructed in vitro by assembling the structural and functional elements that contribute to a complete chromosome capable of stable replication and segregation alongside endogenous chromosomes in cells. The identification of the discrete elements that in combination yield a functional chromosome has made
- 10 possible the in vitro generation of artificial chromosomes. The process of in vitro construction of artificial chromosomes, which can be rigidly controlled, provides advantages that may be desired in the generation of chromosomes that, for example, are required in large amounts or that are intended for specific use in transgenic animal systems.
- 15 For example, in vitro construction may be advantageous when efficiency of time and scale are important considerations in the preparation of artificial chromosomes. Because in vitro construction methods do not involve extensive cell culture procedures, they may be utilized when the time and labor required to transform, feed, cultivate,
- 20 and harvest cells used in in vivo cell-based production systems is unavailable.

- In vitro construction may also be rigorously controlled with respect to the exact manner in which the several elements of the desired artificial chromosome are combined and in what sequence and proportions they
- 25 are assembled to yield a chromosome of precise specifications. These aspects may be of significance in the production of artificial chromosomes that will be used in live animals where it is desirable to be certain that only very pure and specific DNA sequences in specific amounts are being introduced into the host animal.

The following describes the processes involved in the construction of artificial chromosomes in vitro, utilizing a megachromosome as exemplary starting material.

1. Identification and isolation of the components of the artificial chromosome

- 5 The MACs provided herein, particularly the SATACs, are elegantly simple chromosomes for use in the identification and isolation of components to be used in the in vitro construction of artificial chromosomes. The ability to purify MACs to a very high level of purity,
- 10 as described herein, facilitates their use for these purposes. For example, the megachromosome, particularly truncated forms thereof [i.e. cell lines, such as 1B3 and mM2C1, which are derived from H1D3 (deposited at the European Collection of Animal Cell Culture (ECACC) under Accession No. 96040929, see EXAMPLES below) serve as starting materials.
- 15 For example, the mM2C1 cell line contains a micro-megachromosome (~50-60 kB), which advantageously contains only one centromere, two regions of integrated heterologous DNA with adjacent rDNA sequences, with the remainder of the chromosomal DNA being mouse major satellite DNA. Other truncated megachromosomes can
- 20 serve as a source of telomeres, or telomeres can be provided (see, Examples below regarding construction of plasmids containing tandemly repeated telomeric sequences). The centromere of the mM2C1 cell line contains mouse minor satellite DNA, which provides a useful tag for isolation of the centromeric DNA.
- 25 Additional features of particular SATACs provided herein, such as the micro-megachromosome of the mM2C1 cell line, that make them uniquely suited to serve as starting materials in the isolation and identification of chromosomal components include the fact that the centromeres of each megachromosome within a single specific cell line
- 30 are identical. The ability to begin with a homogeneous centromere source (as opposed to a mixture of different chromosomes having differing

- centromeric sequences) greatly facilitates the cloning of the centromere DNA. By digesting purified megachromosomes, particularly truncated megachromosomes, such as the micro-megachromosome, with appropriate restriction endonucleases and cloning the fragments into the commercially available and well known YAC vectors (see, e.g., Burke et al. (1987) Science 236:806-812), BAC vectors (see, e.g., Shizuya et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89: 8794-8797 bacterial artificial chromosomes which have a capacity of incorporating 0.9 - 1 Mb of DNA) or PAC vectors (the P1 artificial chromosome vector which is a P1 plasmid derivative that has a capacity of incorporating 300 kb of DNA and that is delivered to E. coli host cells by electroporation rather than by bacteriophage packaging; see, e.g., Ioannou et al. (1994) Nature Genetics 6:84-89; Pierce et al. (1992) Meth. Enzymol. 216:549-574; Pierce et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:2056-2060; U.S. Patent No. 5,300,431 and International PCT application No. WO 92/14819) vectors, it is possible for as few as 50 clones to represent the entire micro-megachromosome.

a. Centromeres

- An exemplary centromere for use in the construction of a mammalian artificial chromosome is that contained within the megachromosome of any of the megachromosome-containing cell lines provided herein, such as, for example, H1D3 and derivatives thereof, such as mM2C1 cells. Megachromosomes are isolated from such cell lines utilizing, for example, the procedures described herein, and the centromeric sequence is extracted from the isolated megachromosomes. For example, the megachromosomes may be separated into fragments utilizing selected restriction endonucleases that recognize and cut at sites that, for instance, are primarily located in the replication and/or heterologous DNA integration sites and/or in the satellite DNA. Based on the sizes of the resulting fragments, certain undesired elements may be

separated from the centromere-containing sequences. The centromere-containing DNA, which could be as large as 1 Mb.

- Probes that specifically recognize the centromeric sequences, such as mouse minor satellite DNA-based probes [see, e.g., Wong *et al.* (1988) *Nucl. Acids Res.* **16**:11645-11661], may be used to isolate the centromere-containing YAC, BAC or PAC clones derived from the megachromosome. Alternatively, or in conjunction with the direct identification of centromere-containing megachromosomal DNA, probes that specifically recognize the non-centromeric elements, such as probes specific for mouse major satellite DNA, the heterologous DNA and/or rDNA, may be used to identify and eliminate the non-centromeric DNA-containing clones.

- Additionally, centromere cloning methods described herein may be utilized to isolate the centromere-containing sequence of the megachromosome. For example, Example 12 describes the use of YAC vectors in combination with the murine tyrosinase gene and NMRI/Han mice for identification of the centromeric sequence.

- Once the centromere fragment has been isolated, it may be sequenced and the sequence information may in turn be used in PCR amplification of centromere sequences from megachromosomes or other sources of centromeres. Isolated centromeres may also be tested for function *in vivo* by transferring the DNA into a host mammalian cell. Functional analysis may include, for example, examining the ability of the centromere sequence to bind centromere-binding proteins. The cloned centromere will be transferred to mammalian cells with a selectable marker gene and the binding of a centromere-specific protein, such as anti-centromere antibodies (e.g., LU851, see, Hadlaczy *et al.* (1986) *Exp. Cell Res.* **167**:1-15) can be used to assess function of the centromeres.

b. Telomeres

Preferred telomeres are the 1 kB synthetic telomere provided herein (see, Examples). A double synthetic telomere construct, which contains a 1 kB synthetic telomere linked to a dominant selectable marker gene that continues in an inverted orientation may be used for ease of manipulation. Such a double construct contains a series of TTAGGG repeats 3' of the marker gene and a series of repeats of the inverted sequence, i.e., GGGATT, 5' of the marker gene as follows:
 (GGGATTT)_n---dominant marker gene---(TTAGGG)_n. Using an inverted marker provides an easy means for insertion, such as by blunt end ligation, since only properly oriented fragments will be selected.

c. Megareplicator

The megareplicator sequences, such as the rDNA, provided herein are preferred for use in in vitro constructs. The rDNA provides an origin of replication and also provides sequences that facilitate amplification of the artificial chromosome in vivo to increase the size of the chromosome to, for example accommodate increasing copies of a heterologous gene of interest as well as continuous high levels of expression of the heterologous genes.

d. Filler heterochromatin

Filler heterochromatin, particularly satellite DNA, is included to maintain structural integrity and stability of the artificial chromosome and provide a structural base for carrying genes within the chromosome. The satellite DNA is typically A/T-rich DNA sequence, such as mouse major satellite DNA, or G/C-rich DNA sequence, such as hamster natural satellite DNA. Sources of such DNA include any eukaryotic organisms that carry non-coding satellite DNA with sufficient A/T or G/C composition to promote ready separation by sequence, such as by FACS, or by density gradients. The satellite DNA may also be synthesized by generating sequence containing monotone, tandem repeats of highly A/T- or G/C-rich DNA units.

- The most suitable amount of filler heterochromatin for use in construction of the artificial chromosome may be empirically determined by, for example, including segments of various lengths, increasing in size, in the construction process. Fragments that are too small to be suitable
- 5 for use will not provide for a functional chromosome, which may be evaluated in cell-based expression studies, or will result in a chromosome of limited functional lifetime or mitotic and structural stability.

e. Selectable marker

- Any convenient selectable marker may be used and at any
- 10 convenient locus in the MAC.

2. Combination of the isolated chromosomal elements

- Once the isolated elements are obtained, they may be combined to generate the complete, functional artificial chromosome. This assembly can be accomplished for example, by in vitro ligation either in solution,
- 15 LMP agarose or on microbeads. The ligation is conducted so that one end of the centromere is directly joined to a telomere. The other end of the centromere, which serves as the gene-carrying chromosome arm, is built up from a combination of satellite DNA and rDNA sequence and may also contain a selectable marker gene. Another telomere is joined to the end
- 20 of the gene-carrying chromosome arm. The gene-carrying arm is the site at which any heterologous genes of interest, for example, in expression of desired proteins encoded thereby, are incorporated either during in vitro construction of the chromosome or sometime thereafter.

3. Analysis and testing of the artificial chromosome

- 25 Artificial chromosomes constructed in vitro may be tested for functionality in in vivo mammalian cell systems, using any of the methods described herein for the SATACs, minichromosomes, or known to those of skill in the art.

4. Introduction of desired heterologous DNA into the *in vitro* synthesized chromosome

- Heterologous DNA may be introduced into the *in vitro* synthesized chromosome using routine methods of molecular biology, may be introduced using the methods described herein for the SATACs, or may be incorporated into the *in vitro* synthesized chromosome as part of one of the synthetic elements, such as the heterochromatin. The heterologous DNA may be linked to a selected repeated fragment, and then the resulting construct may be amplified *in vitro* using the methods for such *in vitro* amplification provided herein (see the Examples).

D. Introduction of artificial chromosomes into cells, tissues, animals and plants

- Suitable hosts for introduction of the MACs provided herein, include, but are not limited to, any animal or plant, cell or tissue thereof, including, but not limited to: mammals, birds, reptiles, amphibians, insects, fish, arachnids, tobacco, tomato, wheat, plants and algae. The MACs, if contained in cells, may be introduced by cell fusion or microcell fusion or, if the MACs have been isolated from cells, they may be introduced into host cells by any method known to those of skill in this art, including but not limited to: direct DNA transfer, electroporation, lipid-mediated transfer, *e.g.*, lipofection and liposomes, microprojectile bombardment, microinjection in cells and embryos, protoplast regeneration for plants, and any other suitable method [see, *e.g.*, Weissbach *et al.* (1988) *Methods for Plant Molecular Biology*, Academic Press, N.Y., Section VIII, pp. 421-463; Grierson *et al.* (1988) *Plant Molecular Biology*, 2d Ed., Blackie, London, Ch. 7-9; see, also U.S. Patent Nos. 5,491,075; 5,482,928; and 5,424,409; see, also, *e.g.*, U.S. Patent No. 5,470,708, which describes particle-mediated transformation of mammalian unattached cells].
- Other methods for introducing DNA into cells include nuclear microinjection and bacterial protoplast fusion with intact cells. Polycations, such as polybrene and polyornithine, may also be used. For

various techniques for transforming mammalian cells, see e.g., Keown et al. Methods in Enzymology (1990) Vol. 185, pp. 527-537; and Mansour et al. (1988) Nature 336:348-352.

- For example, isolated, purified artificial chromosomes can be
- 5 injected into an embryonic cell line such as a human kidney primary embryonic cell line [ATCC accession number CRL 1573] or embryonic stem cells [see, e.g., Hogan et al. (1994) *Manipulating the Mouse Embryo, A :Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, see, especially, pages 255-264 and Appendix 3].
- 10 Preferably the chromosomes are introduced by microinjection, using a system such as the Eppendorf automated microinjection system, and grown under selective conditions, such as in the presence of hygromycin B or neomycin.

1. Methods for introduction of chromosomes into hosts

- 15 Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. These methods include any, including those described herein, known to those of skill in the art.

a. DNA uptake

- For mammalian cells that do not have cell walls, the calcium
- 20 phosphate precipitation method for introduction of exogenous DNA [see, e.g., Graham et al. (1978) Virology 52:456-457; Wigler et al. (1979) Proc. Natl. Acad. Sci. U.S.A., 76:1373-1376; and Current Protocols in Molecular Biology, Vol. 1, Wiley Inter-Science, Supplement 14, Unit 9.1.1-9.1.9 (1990)] is often preferred. DNA uptake can be accomplished
- 25 by DNA alone or in the presence of polyethylene glycol [PEG-mediated gene transfer], which is a fusion agent, or by any variations of such methods known to those of skill in the art [see, e.g., U.S. Pat. No. 4,684,611].

- Lipid-mediated carrier systems are also among the preferred
- 30 methods for introduction of DNA into cells [see, e.g., Teifel et al. (1995) Biotechniques 19:79-80; Albrecht et al. (1996) Ann. Hematol. 72:73-79;

Holmen *et al.* (1995) *In Vitro Cell Dev. Biol. Anim.* 31:347-351; Remy *et al.* (1994) *Bioconjug. Chem.* 5:647-654; Le Bolc'h *et al.* (1995) *Tetrahedron Lett.* 36:6681-6684; Loeffler *et al.* (1993) *Meth. Enzymol.* 217:599-618]. Lipofection [see, *e.g.*, Strauss (1996) *Meth. Mol. Biol.* 54:307-327] may also be used to introduce DNA into cells. This method is particularly well-suited for transfer of exogenous DNA into chicken cells (*e.g.*, chicken blastodermal cells and primary chicken fibroblasts; see Brazolot *et al.* (1991) *Mol. Repro. Dev.* 30:304-312). In particular, DNA of interest can be introduced into chickens in operative linkage with promoters from genes, such as lysozyme and ovalbumin, that are expressed in the egg, thereby permitting expression of the heterologous DNA in the egg.

Additional methods useful in the direct transfer of DNA into cells include particle gun electrofusion [see, *e.g.*, U.S. Patent Nos. 4,955,378, 4,923,814, 4,476,004, 4,906,576 and 4,441,972] and virion-mediated gene transfer.

A commonly used approach for gene transfer in land plants involves the direct introduction of purified DNA into protoplasts. The three basic methods for direct gene transfer into plant cells include: 1) polyethylene glycol (PEG)-mediated DNA uptake, 2) electroporation-mediated DNA uptake and 3) microinjection. In addition, plants may be transformed using ultrasound treatment [see, *e.g.*, International PCT application publication No. WO 91/00358].

b. Electroporation

Electroporation involves providing high-voltage electrical pulses to a solution containing a mixture of protoplasts and foreign DNA to create reversible pores in the membranes of plant protoplasts as well as other cells. Electroporation is generally used for prokaryotes or other cells, such as plants that contain substantial cell-wall barriers. Methods for effecting electroporation are well known [see, *e.g.*, U.S. Patent Nos.

4,784,737, 5,501,967, 5,501,662, 5,019,034, 5,503,999; see, also Frommet et al. (1985) Proc. Natl. Acad. Sci. U.S.A. 82:5824-5828].

For example, electroporation is often used for transformation of plants [see, e.g., Ag Biotechnology News 7:3 and 17 (September/October 1990)]. In this technique, plant protoplasts are electroporated in the presence of the DNA of interest that also includes a phenotypic marker. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

- 10 Transformed plant cells will be identified by virtue of the expressed phenotypic marker. The exogenous DNA may be added to the protoplasts in any form such as, for example, naked linear, circular or supercoiled DNA, DNA encapsulated in liposomes, DNA in spheroplasts, DNA in other plant protoplasts, DNA complexed with salts, and other methods.

15 c. Microcells

The chromosomes can be transferred by preparing microcells containing an artificial chromosome and then fusing with selected target cells. Methods for such preparation and fusion of microcells are well known [see the Examples and also see, e.g., U.S. Patent Nos. 5,240,840, 20 4,806,476, 5,298,429, 5,396,767, Fournier (1981) Proc. Natl. Acad. Sci. U.S.A. 78:6349-6353; and Lambert et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:5907-59]. Microcell fusion, using microcells that contain an artificial chromosome, is a particularly useful method for introduction of MACs into avian cells, such as DT40 chicken pre-B cells [for a description of DT40 cell fusion, see, e.g., Dieken et al. (1996) Nature Genet. 12:174-182].

2. Hosts

- Suitable hosts include any host known to be useful for introduction and expression of heterologous DNA. Of particular interest herein, animal 30 and plant cells and tissues, including, but not limited to insect cells and larvae, plants, and animals, particularly transgenic (non-human) animals,

- and animal cells. Other hosts include, but are not limited to mammals, birds, particularly fowl such as chickens, reptiles, amphibians, insects, fish, arachnids, tobacco, tomato, wheat, monocots, dicots and algae, and any host into which introduction of heterologous DNA is desired. Such
- 5 introduction can be effected using the MACs provided herein, or, if necessary by using the MACs provided herein to identify species-specific centromeres and/or functional chromosomal units and then using the resulting centromeres or chromosomal units as artificial chromosomes, or alternatively, using the methods exemplified herein for production of
- 10 MACs to produce species-specific artificial chromosomes.

a. **Introduction of DNA into embryos for production of transgenic (non-human) animals and introduction of DNA into animal cells**

- Transgenic (non-human) animals can be produced by introducing
- 15 exogenous genetic material into a pronucleus of a mammalian zygote by microinjection [see, e.g., U.S. Patent Nos. 4,873,191 and 5,354,674; see, also, International PCT application publication No. WO 95/14769, which is based on U.S. application Serial No. 08/159,084]. The zygote is capable of development into a mammal. The embryo or zygote is
- 20 transplanted into a host female uterus and allowed to develop. Detailed protocols and examples are set forth below.

- Nuclear transfer [see, Wilmut et al. (1997) Nature 385:810-813, International PCT application Nos. WO 97/07669 and WO 97/07668]. Briefly in this method, the SATAC containing the genes of interest is
- 25 introduced by any suitable method, into an appropriate donor cell, such as a mammary gland cell, that contains totipotent nuclei. The diploid nucleus of the cell, which is either in G0 or G1 phase, is then introduced, such as by cell fusion or microinjection, into an unactivated oocyte, preferably enucleated cell, which is arrested in the metaphase of the
- 30 second meiotic division. Enucleation may be effected by any suitable method, such as actual removal, or by treating with means, such as ultraviolet light, that functionally remove the nucleus. The oocyte is then

- activated, preferably after a period of contact, about 6-20 hours for cattle, of the new nucleus with the cytoplasm, while maintaining correct ploidy, to produce a reconstituted embryo, which is then introduced into a host. Ploidy is maintained during activation, for example, by incubating
- 5 the reconstituted cell in the presence of a microtubule inhibitor, such as nocodazole, colchicine, cocemid, and taxol, whereby the DNA replicates once.

- Transgenic chickens can be produced by injection of dispersed blastodermal cells from Stage X chicken embryos into recipient embryos
- 10 at a similar stage of development [see e.g., Etches et al. (1993) Poultry Sci. 72:882-889; Petitte et al. (1990) Development 108:185-189]. Heterologous DNA is first introduced into the donor blastodermal cells using methods such as, for example, lipofection [see, e.g., Brazolot et al. (1991) Mol. Repro. Dev. 30:304-312] or microcell fusion [see, e.g.,
- 15 Dieken et al. (1996) Nature Genet. 12:174-182]. The transfected donor cells are then injected into recipient chicken embryos [see e.g., Carsience et al. (1993) Development 117: 669-675]. The recipient chicken embryos within the shell are candled and allowed to hatch to yield a germline chimeric chicken.
- 20 DNA can be introduced into animal cells using any known procedure, including, but not limited to: direct uptake, incubation with polyethylene glycol [PEG], microinjection, electroporation, lipofection, cell fusion, microcell fusion, particle bombardment, including microprojectile bombardment [see, e.g., U.S. Patent No. 5,470,708, which provides a
- 25 method for transforming unattached mammalian cells via particle bombardment], and any other such method. For example, the transfer of plasmid DNA in liposomes directly to human cells *in situ* has been approved by the FDA for use in humans [see, e.g., Nabel, et al. (1990) Science 249:1285-1288 and U.S. Patent No. 5,461,032].

b. Introduction of heterologous DNA into plants

Numerous methods for producing or developing transgenic plants are available to those of skill in the art. The method used is primarily a function of the species of plant. These methods include, but are not

- 5 limited to: direct transfer of DNA by processes, such as PEG-induced DNA uptake, protoplast fusion, microinjection, electroporation, and microprojectile bombardment [see, e.g., Uchimiya et al. (1989) J. of Biotech. 12: 1-20 for a review of such procedures, see, also, e.g., U.S. Patent Nos. 5,436,392 and 5,489,520 and many others]. For purposes
10 herein, when introducing a MAC, microinjection, protoplast fusion and particle gun bombardment are preferred.

- Plant species, including tobacco, rice, maize, rye, soybean, Brassica napus, cotton, lettuce, potato and tomato, have been used to produce transgenic plants. Tobacco and other species, such as petunias,
15 often serve as experimental models in which the methods have been developed and the genes first introduced and expressed.

- DNA uptake can be accomplished by DNA alone or in the presence of PEG, which is a fusion agent, with plant protoplasts or by any variations of such methods known to those of skill in the art [see, e.g.,
20 U.S. Patent No. 4,684,611 to Schilperoot et al.]. Electroporation, which involves high-voltage electrical pulses to a solution containing a mixture of protoplasts and foreign DNA to create reversible pores, has been used, for example, to successfully introduce foreign genes into rice and Brassica napus. Microinjection of DNA into plant cells, including cultured cells and
25 cells in intact plant organs and embryoids in tissue culture and microprojectile bombardment [acceleration of small high density particles, which contain the DNA, to high velocity with a particle gun apparatus, which forces the particles to penetrate plant cell walls and membranes] have also been used. All plant cells into which DNA can be introduced
30 and that can be regenerated from the transformed cells can be used to produce transformed whole plants which contain the transferred artificial

chromosome. The particular protocol and means for introduction of the DNA into the plant host may need to be adapted or refined to suit the particular plant species or cultivar.

c. Insect cells

- 5 Insects are useful hosts for introduction of artificial chromosomes for numerous reasons, including, but not limited to: (a) amplification of genes encoding useful proteins can be accomplished in the artificial chromosome to obtain higher protein yields in insect cells; (b) insect cells support required post-translational modifications, such as glycosylation
- 10 and phosphorylation, that can be required for protein biological functioning; (c) insect cells do not support mammalian viruses, and, thus, eliminate the problem of cross-contamination of products with such infectious agents; (d) this technology circumvents traditional recombinant baculovirus systems for production of nutritional, industrial or medicinal
- 15 proteins in insect cell systems; (e) the low temperature optimum for insect cell growth (28° C) permits reduced energy cost of production; (f) serum-free growth medium for insect cells permits lower production costs; (g) artificial chromosome-containing cells can be stored indefinitely at low temperature; and (h) insect larvae will be biological factories for
- 20 production of nutritional, medicinal or industrial proteins by microinjection of fertilized insect eggs [see, e.g., Joy et al. (1991) Current Science 66:145-150, which provides a method for microinjecting heterologous DNA into *Bombyx mori* eggs].

- Either MACs or insect-specific artificial chromosomes [BUGACs]
- 25 will be used to introduce genes into insects. As described in the Examples, it appears that MACs will function in insects to direct expression of heterologous DNA contained thereon. For example, as described in the Examples, a MAC containing the *B. mori* actin gene promoter fused to the lacZ gene has been generated by transfection of
- 30 EC3/7C5 cells with a plasmid containing the fusion gene. Subsequent fusion of the *B. mori* cells with the transfected EC3/7C5 cells that

survived selection yielded a MAC-containing insect-mouse hybrid cell line in which β -galactosidase expression was detectable.

- Insect host cells include, but are not limited to, hosts such as *Spodoptera frugiperda* [caterpillar], *Aedes aegypti* [mosquito], *Aedes albopictus* [mosquito], *Drosophila melanogaster* [fruitfly], *Bombyx mori* [silkworm], *Manduca sexta* [tomato horn worm] and *Trichoplusia ni* [cabbage looper]. Efforts have been directed toward propagation of insect cells in culture. Such efforts have focused on the fall armyworm, *Spodoptera frugiperda*. Cell lines have been developed also from other insects such as the cabbage looper, *Trichoplusia ni* and the silkworm, *Bombyx mori*. It has also been suggested that analogous cell lines can be created using the tomato hornworm, *Manduca sexta*. To introduce DNA into an insect, it should be introduced into the larvae, and allowed to proliferate, and then the hemolymph recovered from the larvae so that the proteins can be isolated therefrom.

The preferred method herein for introduction of artificial chromosomes into insect cells is microinjection [see, e.g., Tamura et al. (1991) Bio Ind. 8:26-31; Nikolaev et al. (1989) Mol. Biol. (Moscow) 23:1177-87; and methods exemplified and discussed herein].

20 E. Applications for and Uses of Artificial chromosomes

- Artificial chromosomes provide convenient and useful vectors, and in some instances [e.g., in the case of very large heterologous genes] the only vectors, for introduction of heterologous genes into hosts. Virtually any gene of interest is amenable to introduction into a host via artificial chromosomes. Such genes include, but are not limited to, genes that encode receptors, cytokines, enzymes, proteases, hormones, growth factors, antibodies, tumor suppressor genes, therapeutic products and multigene pathways.

The artificial chromosomes provided herein will be used in methods of protein and gene product production, particularly using insects as host cells for production of such products, and in cellular (e.g., mammalian cell) production systems in which the artificial chromomsomes

- 5 (particularly MACs) provide a reliable, stable and efficient means for optimizing the biomanufacturing of important compounds for medicine and industry. They are also intended for use in methods of gene therapy, and for production of transgenic plants and animals [discussed above, below and in the EXAMPLES].

10 **1. Gene Therapy**

Any nucleic acid encoding a therapeutic gene product or product of a multigene pathway may be introduced into a host animal, such as a human, or into a target cell line for introduction into an animal, for therapeutic purposes. Such therapeutic purposes include, genetic therapy

- 15 to cure or to provide gene products that are missing or defective, to deliver agents, such as anti-tumor agents, to targeted cells or to an animal, and to provide gene products that will confer resistance or reduce susceptibility to a pathogen or ameliorate symptoms of a disease or disorder. The following are some exemplary genes and gene products.
- 20 Such exemplification is not intended to be limiting.

a. Anti-HIV ribozymes

As exemplified below, DNA encoding anti-HIV ribozymes can be introduced and expressed in cells using MACs, including the euchromatin-based minichromosomes and the SATACs. These MACs can be used to

- 25 make a transgenic mouse that expresses a ribozyme and, thus, serves as a model for testing the activity of such ribozymes or from which ribozyme-producing cell lines can be made. Also, introduction of a MAC that encodes an anti-HIV ribozyme into human cells will serve as

treatment for HIV infection. Such systems further demonstrate the viability of using any disease-specific ribozyme to treat or ameliorate a particular disease.

b. Tumor Suppressor Genes

5 Tumor suppressor genes are genes that, in their wild-type alleles, express proteins that suppress abnormal cellular proliferation. When the gene coding for a tumor suppressor protein is mutated or deleted, the resulting mutant protein or the complete lack of tumor suppressor protein expression may result in a failure to correctly regulate cellular

10 proliferation. Consequently, abnormal cellular proliferation may take place, particularly if there is already existing damage to the cellular regulatory mechanism. A number of well-studied human tumors and tumor cell lines have been shown to have missing or nonfunctional tumor suppressor genes.

15 Examples of tumor suppression genes include, but are not limited to, the retinoblastoma susceptibility gene or RB gene, the p53 gene, the gene that is deleted in colon carcinoma [*i.e.*, the DCC gene] and the neurofibromatosis type 1 [NF-1] tumor suppressor gene [see, *e.g.*, U.S. Patent No. 5,496,731; Weinberg *et al.* (1991) *254*:1138-1146]. Loss of

20 function or inactivation of tumor suppressor genes may play a central role in the initiation and/or progression of a significant number of human cancers.

The p53 Gene

Somatic cell mutations of the p53 gene are said to be the most

25 frequent of the gene mutations associated with human cancer [see, *e.g.*, Weinberg *et al.* (1991) *Science 254*:1138-1146]. The normal or wild-type p53 gene is a negative regulator of cell growth, which, when damaged, favors cell transformation. The p53 expression product is found in the nucleus, where it may act in parallel or cooperatively with

30 other gene products. Tumor cell lines in which p53 has been deleted

have been successfully treated with wild-type p53 vector to reduce tumorigenicity [see, Baker *et al.* (1990) *Science* 249:912-915].

DNA encoding the p53 gene and plasmids containing this DNA are well known [see, *e.g.*, U.S. Patent No. 5,260,191; see, also Chen *et al.*

- 5 (1990) *Science* 250:1576; Farrel *et al.* (1991) *EMBO J.* 10:2879-2887; plasmids containing the gene are available from the ATCC, and the sequence is in the GenBank Database, accession nos. X54156, X60020, M14695, M16494, K03199].

c. The CFTR gene

- 10 Cystic fibrosis [CF] is an autosomal recessive disease that affects epithelia of the airways, sweat glands, pancreas, and other organs. It is a lethal genetic disease associated with a defect in chloride ion transport, and is caused by mutations in the gene coding for the cystic fibrosis transmembrane conductance regulator [CFTR], a 1480 amino acid protein
- 15 that has been associated with the expression of chloride conductance in a variety of eukaryotic cell types. Defects in CFTR destroy or reduce the ability of epithelial cells in the airways, sweat glands, pancreas and other tissues to transport chloride ions in response to cAMP-mediated agonists and impair activation of apical membrane channels by cAMP-dependent
- 20 protein kinase A [PKA]. Given the high incidence and devastating nature of this disease, development of effective CF treatments is imperative.

- The CFTR gene [~250 kb] can be transferred into a MAC for use, for example, in gene therapy as follows. A CF-YAC [see Green *et al.* *Science* 250:94-98] may be modified to include a selectable marker, such
- 25 as a gene encoding a protein that confers resistance to puromycin or hygromycin, and λ -DNA for use in site-specific integration into a neo-minichromosome or a SATAC. Such a modified CF-YAC can be introduced into MAC-containing cells, such as EC3/7C5 or 19C5xHa4 cells, by fusion with yeast protoplasts harboring the modified CF-YAC or
- 30 microinjection of yeast nuclei harboring the modified CF-YAC into the cells. Stable transformants are then selected on the basis of antibiotic

resistance. These transformants will carry the modified CF-YAC within the MAC contained in the cells.

2. Animals, birds, fish and plants that are genetically altered to possess desired traits such as resistance to disease

5 Artificial chromosomes are ideally suited for preparing animals, including vertebrates and invertebrates, including birds and fish as well as mammals, that possess certain desired traits, such as, for example, disease resistance, resistance to harsh environmental conditions, altered growth patterns, and enhanced physical characteristics.

10 One example of the use of artificial chromosomes in generating disease-resistant organisms involves the preparation of multivalent vaccines. Such vaccines include genes encoding multiple antigens that can be carried in a MAC, or species-specific artificial chromosome, and either delivered to a host to induce immunity, or incorporated into
15 embryos to produce transgenic (non-human) animals and plants that are immune or less susceptible to certain diseases.

Disease-resistant animals and plants may also be prepared in which resistance or decreased susceptibility to disease is conferred by introduction into the host organism or embryo of artificial chromosomes
20 containing DNA encoding gene products (e.g., ribozymes and proteins that are toxic to certain pathogens) that destroy or attenuate pathogens or limit access of pathogens to the host.

Animals and plants possessing desired traits that might, for example, enhance utility, processibility and commercial value of the
25 organisms in areas such as the agricultural and ornamental plant industries may also be generated using artificial chromosomes in the same manner as described above for production of disease-resistant animals and plants. In such instances, the artificial chromosomes that are introduced into the organism or embryo contain DNA encoding gene
30 products that serve to confer the desired trait in the organism.

Birds, particularly fowl such as chickens, fish and crustaceans will serve as model hosts for production of genetically altered organisms using artificial chromosomes.

5 3. Use of MACs and other artificial chromosomes for preparation and screening of libraries

Since large fragments of DNA can be incorporated into each artificial chromosome, the chromosomes are well-suited for use as cloning vehicles that can accommodate entire genomes in the preparation of genomic DNA libraries, which then can be readily screened. For example,

10 MACs may be used to prepare a genomic DNA library useful in the identification and isolation of functional centromeric DNA from different species of organisms. In such applications, the MAC used to prepare a genomic DNA library from a particular organism is one that is not functional in cells of that organism. That is, the MAC does not stably

15 replicate, segregate or provide for expression of genes contained within it in cells of the organism. Preferably, the MACs contain an indicator gene (e.g., the lacZ gene encoding β -galactosidase or genes encoding products that confer resistance to antibiotics such as neomycin, puromycin, hygromycin) linked to a promoter that is capable of promoting

20 transcription of the indicator gene in cells of the organism. Fragments of genomic DNA from the organism are incorporated into the MACs, and the MACs are transferred to cells from the organism. Cells that contain MACs that have incorporated functional centromeres contained within the genomic DNA fragments are identified by detection of expression of the

25 marker gene.

4. Use of MACs and other artificial chromosomes for stable, high-level protein production

Cells containing the MACs and/or other artificial chromosomes provided herein are advantageously used for production of proteins,

30 particularly several proteins from one cell line, such as multiple proteins involved in a biochemical pathway or multivalent vaccines. The genes encoding the proteins are introduced into the artificial chromosomes

which are then introduced into cells. Alternatively, the heterologous gene(s) of interest are transferred into a production cell line that already contains artificial chromosomes in a manner that targets the gene(s) to the artificial chromosomes. The cells are cultured under conditions

- 5 whereby the heterologous proteins are expressed. Because the proteins will be expressed at high levels in a stable permanent extra-genomic chromosomal system, selective conditions are not required.

- Any transfectable cells capable of serving as recombinant hosts adaptable to continuous propagation in a cell culture system [see, e.g.,
 10 McLean (1993) Trends In Biotech. 11:232-238] are suitable for use in an artificial chromosome-based protein production system. Exemplary host cell lines include, but are not limited to, the following: Chinese hamster ovary (CHO) cells [see, e.g., Zang et al. (1995) Biotechnology 13:389-392], HEK 293, Ltk⁻, COS-7, DG44, and BHK cells. CHO cells are
 15 particularly preferred host cells. Selection of host cell lines for use in artificial chromosome-based protein production systems is within the skill of the art, but often will depend on a variety of factors, including the properties of the heterologous protein to be produced, potential toxicity of the protein in the host cell, any requirements for post-translational
 20 modification (e.g., glycosylation, amination, phosphorylation) of the protein, transcription factors available in the cells, the type of promoter element(s) being used to drive expression of the heterologous gene, whether production will be completely intracellular or the heterologous protein will preferably be secreted from the cell, and the types of
 25 processing enzymes in the cell.

- The artificial chromosome-based system for heterologous protein production has many advantageous features. For example, as described above, because the heterologous DNA is located in an independent, extra-genomic artificial chromosome (as opposed to randomly inserted in an
 30 unknown area of the host cell genome or located as extrachromosomal element(s) providing only transient expression) it is stably maintained in

- an active transcription unit and is not subject to ejection via recombination or elimination during cell division. Accordingly, it is unnecessary to include a selection gene in the host cells and thus growth under selective conditions is also unnecessary. Furthermore, because the
- 5 artificial chromosomes are capable of incorporating large segments of DNA, multiple copies of the heterologous gene and linked promoter element(s) can be retained in the chromosomes, thereby providing for high-level expression of the foreign protein(s). Alternatively, multiple
 - 10 different genes may be linked in a fused polygene complex to a single promoter for expression of, for example, all the key proteins constituting a complete metabolic pathway [see, e.g., Beck von Bodman et al. (1995) Biotechnology 13:587-591]. Alternatively, multiple copies of a single gene can be operatively linked to a single promoter, or each or one or several
 - 15 copies may be linked to different promoters or multiple copies of the same promoter. Additionally, because artificial chromosomes have an almost unlimited capacity for integration and expression of foreign genes, they can be used not only for the expression of genes encoding end-products of interest, but also for the expression of genes associated with optimal
 - 20 maintenance and metabolic management of the host cell, e.g., genes encoding growth factors, as well as genes that may facilitate rapid synthesis of correct form of the desired heterologous protein product, e.g., genes encoding processing enzymes and transcription factors. The MACS are suitable for expression of any proteins or peptides,
 - 25 including proteins and peptides that require in vivo posttranslational modification for their biological activity. Such proteins include, but are not limited to antibody fragments, full-length antibodies, and multimeric antibodies, tumor suppressor proteins, naturally occurring or artificial antibodies and enzymes, heat shock proteins, and others.
 - 30 Thus, such cell-based "protein factories" employing MACs can generated using MACs constructed with multiple copies [theoretically an

- unlimited number or at least up to a number such that the resulting MAC is about up to the size of a genomic chromosome (i.e., endogenous)] of protein-encoding genes with appropriate promoters, or multiple genes driven by a single promoter, i.e., a fused gene complex [such as a
- 5 complete metabolic pathway in plant expression system; see, e.g., Beck von Bodman (1995) Biotechnology 13:587-591]. Once such MAC is constructed, it can be transferred to a suitable cell culture system, such as a CHO cell line in protein-free culture medium [see, e.g., (1995) Biotechnology 13:389-39] or other immortalized cell lines [see, e.g.,
- 10 (1993) TIBTECH 11:232-238] where continuous production can be established.

- The ability of MACs to provide for high-level expression of heterologous proteins in host cells is demonstrated, for example, by analysis of the H1D3 and G3D5 cell lines described herein and deposited
- 15 with the ECACC. Northern blot analysis of mRNA obtained from these cells reveals that expression of the hygromycin-resistance and β -galactosidase genes in the cells correlates with the amplicon number of the megachromosome(s) contained therein.

20 **F. Methods for the synthesis of DNA sequences containing repeated DNA units**

- Generally, assembly of tandemly repeated DNA poses difficulties such as unambiguous annealing of the complementary oligos. For example, separately annealed products may ligate in an inverted
- 25 orientation. Additionally, tandem or inverted repeats are particularly susceptible to recombination and deletion events that may disrupt the sequence. Selection of appropriate host organisms (e.g., rec^- strains) for use in the cloning steps of the synthesis of sequences of tandemly repeated DNA units may aid in reduction and elimination of such events.

- 30 Methods are provided herein for the synthesis of extended DNA sequences containing repeated DNA units. These methods are particularly applicable to the synthesis of arrays of tandemly repeated

DNA units, which are generally difficult or not possible to construct utilizing other known gene assembly strategies. A specific use of these methods is in the synthesis of sequences of any length containing simple (e.g., ranging from 2-6 nucleotides) tandem repeats (such as telomeres and satellite DNA repeats and trinucleotide repeats of possible clinical significance) as well as complex repeated DNA sequences. An particular example of the synthesis of a telomere sequence containing over 150 successive repeated hexamers utilizing these methods is provided herein.

- The methods provided herein for synthesis of arrays of tandem
- 10 DNA repeats are based in a series of extension steps in which successive doublings of a sequence of repeats results in an exponential expansion of the array of tandem repeats. These methods provide several advantages over previously known methods of gene assembly. For instance, the starting oligonucleotides are used only once. The intermediates in, as
- 15 well as the final product of, the construction of the DNA arrays described herein may be obtained in cloned form in a microbial organism (e.g., E. coli and yeast). Of particular significance, with regard to these methods is the fact that sequence length increases exponentially, as opposed to linearly, in each extension step of the procedure even though only two
- 20 oligonucleotides are required in the methods. The construction process does not depend on the compatibility of restriction enzyme recognition sequences and the sequence of the repeated DNA because restriction sites are used only temporarily during the assembly procedure. No adaptor is necessary, though a region of similar function is located
- 25 between two of the restriction sites employed in the process. The only limitation with respect to restriction site use is that the two sites employed in the method must not be present elsewhere in the vector utilized in any cloning steps. These procedures can also be used to construct complex repeats with perfectly identical repeat units, such as
- 30 the variable number tandem repeat (VNTR) 3' of the human apolipoprotein B100 gene (a repeat unit of 30 bp, 100% AT) or aliphoid satellite DNA.

[illegible][illegible][illegible]

- [illegible]

[illegible]

- $$\begin{array}{ccccccc} \text{atoms} & \text{links} & \text{nodes} & \text{edges} & \text{faces} & \text{cells} & \text{volumes} \\ b_0, d_0 & b_1, d_1 & b_2, d_2 & b_3, d_3 & b_4, d_4 & b_5, d_5 & b_6, d_6 \end{array}$$

$$\begin{array}{ccccccc} \text{atoms} & \text{links} & \text{nodes} & \text{edges} & \text{faces} & \text{cells} & \text{volumes} \\ b_0, d_0 & b_1, d_1 & b_2, d_2 & b_3, d_3 & b_4, d_4 & b_5, d_5 & b_6, d_6 \end{array}$$
[illegible]

- [illegible]

[illegible][illegible]

- [illegible]

Tail: ACGT/ (Klenow/dNTP) --

Additionally, if mung bean nuclease is used instead of Klenow, then the following

XbaI: T/CTAGA Mung bean nuclease A--

10

Furthermore, there are a number of restriction enzymes that cut outside of the recognition sequence, and in this case, there is no limitation at all:

TspRI NNCAGTGNN/-- (Klenow/dNTP) --

BsmI GAATG CN/-- (Klenow/dNTP) --

15 CTTAC/GN -- (Klenow/dNTP) --

2. Step 1 - Annealing

Oligonucleotides 1 and 2 are annealed at a temperature selected depending on the length of overlap (typically in the range of 30-65 °C).

3. Step 2 - Generating a double-stranded molecule

20 The annealed oligonucleotides are filled-in with Klenow polymerase in the presence of dNTP to produce a double-stranded (ds) sequence:

[illegible][illegible]

4. Step 3 - Incorporation of double-stranded DNA into a vector

25 The double-stranded DNA is cleaved with restriction enzymes E1 and E3 and subsequently ligated into a vector (e.g., pUC19 or a yeast vector) that has been cleaved with the same enzymes E1 and E3. The ligation product is used to transform competent host cells compatible with the vector being used (e.g., when pUC19 is used, bacterial cells such as

30 E. coli DH5 α are suitable hosts) which are then plated onto selection plates. Recombinants can be identified either by color (e.g., by X-gal

- staining for β -galactosidase expression) or by colony hybridization using ^{32}P -labeled oligonucleotide 2 (detection by hybridization to oligonucleotide 2 is preferred because its sequence is removed in each of the subsequent extension steps and thus is present only in recombinants that contain
- 5 DNA that has undergone successful extension of the repeated sequence).

5. Step 4 - Isolation of insert from the plasmid

- An aliquot of the recombinant plasmid containing k nucleotides of the repeat sequence is digested with restriction enzymes E1 and E3, and the insert is isolated on a gel (native polyacrylamide while the insert is
- 10 short, but agarose can be used for isolation of longer inserts in subsequent steps). A second aliquot of the recombinant plasmid is cut with enzymes E2 (treated with Klenow and dNTP to remove the 3'-overhang) and E3, and the large fragment (plasmid DNA plus the insert) is isolated.

- 15 **6. Step 5 - Extension of the DNA sequence of k repeats**

- The two DNAs (the S1-S3 insert fragment and the vector plus insert) are ligated, plated to selective plates, and screened for extended recombinants as in Step 3. Now the length of the repeat sequence between restriction sites is twice that of the repeat sequence in the
- 20 previous step, i.e., $2 \times k$.

7. Step 6 - Extension of the DNA sequence of $2 \times k$ repeats

- Steps 4 and 5 are repeated as many times as needed to achieve the desired repeat sequence size. In each extension cycle, the repeat sequence size doubles, i.e., if m is the number of extension cycles, the
- 25 size of the repeat sequence will be $k \times 2^m$ nucleotides.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

General Materials and Methods

The following materials and methods are exemplary of methods that are used in the following Examples and that can be used to prepare cell lines containing artificial chromosomes. Other suitable materials and methods known to those of skill in the art may be used. Modifications of these materials and methods known to those of skill in the art may also be employed.

A. Culture of cell lines, cell fusion, and transfection of cells

1. Chinese hamster K-20 cells and mouse A9 fibroblast cells were cultured in F-12 medium. EC3/7 [see, U.S. Patent No. 5,288,625, and deposited at the European Collection of Animal cell Culture (ECACC) under accession no. 90051001; see, also Hadlaczký *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:8106-8110 and U.S. application Serial No. 08/375,271] and EC3/7C5 [see, U.S. Patent No. 5,288,625 and Praznovszky *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:11042-11046] mouse cell lines, and the KE1-2/4 hybrid cell line were maintained in F-12 medium containing 400 μ g/ml G418 [SIGMA, St. Louis, MO].
2. TF1004G19 and TF1004G-19C5 mouse cells, described below, and the 19C5xHa4 hybrid, described below, and its sublines were cultured in F-12 medium containing up to 400 μ g/ml Hygromycin B [Calbiochem]. LP11 cells were maintained in F-12 medium containing 3-15 μ g/ml Puromycin [SIGMA, St. Louis, MO].
3. Cotransfection of EC3/7C5 cells with plasmids [pH132, pCH110 available from Pharmacia, see, also Hall *et al.* (1983) *J. Mol. Appl. Gen.* 2:101-109] and with λ DNA was conducted using the calcium phosphate DNA precipitation method [see, *e.g.*, Chen *et al.* (1987) *Mol. Cell. Biol.* 7:2745-2752], using 2-5 μ g plasmid DNA and 20 μ g λ phage DNA per 5×10^6 recipient cells.

4. Cell fusion

Mouse and hamster cells were fused using polyethylene glycol [Davidson et al. (1976) Som. Cell Genet. 2:165-176]. Hybrid cells were selected in HAT medium containing 400 μ g/ml Hygromycin B.

- 5 Approximately 2×10^7 recipient and 2×10^6 donor cells were fused using polyethylene glycol [Davidson et al. (1976) Som. Cell Genet. 2:165-176]. Hybrids were selected and maintained in F-12/HAT medium [Szybalsky et al. (1962) Natl. Cancer Inst. Monogr. 7:75-89] containing 10% FCS and 400 μ g/ml G418. The presence of "parental"
- 10 chromosomes in the hybrid cell lines was verified by *in situ* hybridization with species-specific probes using biotin-labeled human and hamster genomic DNA, and a mouse long interspersed repetitive DNA [pMCPE1.51].

5. Microcell fusion

- 15 Microcell-mediated transfer of artificial chromosomes from EC3/7C5 cells to recipient cells was done according to Saxon et al. [(1985) Mol. Cell. Biol. 1:140-146] with the modifications of Goodfellow et al. [(1989) *Techniques for mammalian genome transfer. In Genome Analysis a Practical Approach.* K.E. Davies, ed., IRL Press, Oxford, Washington DC. pp.1-17] and Yamada et al. [(1990) Oncogene 5:1141-1147]. Briefly, 5×10^6 EC3/7C5 cells in a T25 flask were treated first with 0.05 μ g/ml colcemid for 48 hr and then with 10 μ g/ml cytochalasin B for 30 min. The T25 flasks were centrifuged on edge and the pelleted microcells were suspended in serum free DME medium. The microcells were filtered
- 25 through first a 5 micron and then a 3 micron polycarbonate filter, treated with 50 μ g/ml of phytohemagglutinin, and used for polyethylene glycol mediated fusion with recipient cells. Selection of cells containing the MMCneo was started 48 hours after fusion in medium containing 400-800 μ g/ml G418.
- 30 Microcells were also prepared from 1B3 and GHB42 donor cells as follows in order to be fused with E2D6K cells [a CHO K-20 cell line

carrying the puromycin N-acetyltransferase gene, i.e., the puromycin resistance gene, under the control of the SV40 early promoter]. The donor cells were seeded to achieve 60-75% confluency within 24-36 hours. After that time, the cells were arrested in mitosis by exposure to colchicine (10 μ g/ml) for 12 or 24 hours to induce micronucleation. To promote micronucleation of GHB42 cells, the cells were exposed to hypotonic treatment (10 min at 37°C). After colchicine treatment, or after colchicine and hypotonic treatment, the cells were grown in colchicine-free medium.

- 10 The donor cells were trypsinized and centrifuged and the pellets were suspended in a 1:1 Percoll medium and incubated for 30-40 min at 37°C. After the incubation, $1-3 \times 10^7$ cells (60-70% micronucleation index) were loaded onto each Percoll gradient (each fusion was distributed on 1-2 gradients). The gradients were centrifuged at 19,000 rpm for 80 min in a Sorvall SS-34 rotor at 34-37°C. After centrifugation, two visible bands of cells were removed, centrifuged at 2000 rpm, 10 min at 4°C, resuspended and filtered through 8 μ m pore size nucleopore filters.

- 20 The microcells prepared from the 1B3 and GHB42 cells were fused with E2D6K. The E2D6K cells were generated by CaPO_4 transfection of CHO K-20 cells with pCHTV2. Plasmid pCHTV2 contains the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal, the Saccharomyces cerevisiae URA3 gene, 2.4- and 3.2-kb fragments of a Chinese hamster chromosome 2-specific satellite DNA (HC-2 satellite; see Fatyol et al. (1994) Nuc. Acids Res. 22:3728-3736), two copies of the diphtheria toxin-A chain gene (one linked to the herpes simplex virus thymidine kinase (HSV-TK) gene promoter and SV40 polyadenylation signal and the other linked to the HSV-TK promoter without a polyadenylation signal), the ampicillin-resistance gene and the ColE1 origin of replication. Following transfection, puromycin-resistant colonies were isolated. The presence of the pCHTV2 plasmid in the E2D6K cell

line was confirmed by nucleic acid amplification of DNA isolated from the cells.

- The purified microcells were centrifuged as described above and resuspended in 2 ml of phytohemagglutinin-P (PHA-P, 100 μ g/ml). The
- 5 microcell suspension was then added to a 60-70% confluent recipient culture of E2D6K cells. The preparation was incubated at room temperature for 30-40 min to agglutinate the microcells. After the PHA-P was removed, the cells were incubated with 1 ml of 50% polyethylene-glycol (PEG) for one min. The PEG was removed and the culture was
- 10 washed three times with F-12 medium without serum. The cells were incubated in non-selective medium for 48-60 hours. After this time, the cell culture was trypsinized and plated in F-12 medium containing 400 μ g/ml hygromycin B and 10 g/ml puromycin to select against the parental cell lines.
- 15 Hybrid clones were isolated from the cells that had been cultured in selective medium. These clones were then analyzed for expression of β -galactosidase by the X-gal staining method. Four of five hybrid clones analyzed that had been generated by fusion of GHB42 microcells with E2D6K cells yielded positive staining results indicating expression of β -
- 20 galactosidase from the lacZ gene contained in the megachromosome contributed by the GHB42 cells. Similarly, a hybrid clone that had been generated by fusion of 1B3 microcells with E2D6K cells yielded positive staining results indicating expression of β -galactosidase from the lacZ gene contained in the megachromosome contributed by the 1B3 cells. *In*
- 25 *situ* hybridization analysis of the hybrid clones is also performed to analyze the mouse chromosome content of the mouse-hamster hybrid cells.

B. Chromosome banding

- Trypsin G-banding of chromosomes was performed using the
- 30 method of Wang & Fedoroff [(1972) Nature 235:52-54], and the detection of constitutive heterochromatin with the BSG. C-banding

method was done according to Sumner [(1972) Exp. Cell Res. **75**:304-306]. For the detection of chromosome replication by bromodeoxyuridine [BrdU] incorporation, the Fluorescein Plus Giemsa [FPG] staining method of Perry & Wolff [(1974) Nature **251**:156-158] was used.

5 C. Immunolabelling of chromosomes and *in situ* hybridization

Indirect immunofluorescence labelling with human anti-centromere serum LU851 [Hadlaczky et al. (1986) Exp. Cell Res. **167**:1-15], and indirect immunofluorescence and *in situ* hybridization on the same preparation were performed as described previously [see, Hadlaczky et al. **10** (1991) Proc. Natl. Acad. Sci. U.S.A. **88**:8106-8110, see, also U.S. application Serial No. 08/375,271]. Immunolabelling with fluorescein-conjugated anti-BrdU monoclonal antibody [Boehringer] was performed according to the procedure recommended by the manufacturer, except that for treatment of mouse A9 chromosomes, 2 M hydrochloric acid was **15** used at 37° C for 25 min, and for chromosomes of hybrid cells, 1 M hydrochloric acid was used at 37° C for 30 min.

D. Scanning electron microscopy

Preparation of mitotic chromosomes for scanning electron microscopy using osmium impregnation was performed as described **20** previously [Sumner (1991) Chromosoma **100**:410-418]. The chromosomes were observed with a Hitachi S-800 field emission scanning electron microscope operated with an accelerating voltage of 25 kV.

E. DNA manipulations, plasmids and probes

1. General methods

25 All general DNA manipulations were performed by standard procedures [see, e.g., Sambrook et al. (1989) *Molecular cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY]. The mouse major satellite probe was provided by Dr. J. B. Rattner [University of Calgary, Alberta, Canada]. Cloned mouse satellite **30** DNA probes [see Wong et al. (1988) Nucl. Acids Res. **16**:11645-11661], including the mouse major satellite probe, were gifts from Dr. J. B.

Rattner, University of Calgary. Hamster chromosome painting was done with total hamster genomic DNA, and a cloned repetitive sequence specific to the centromeric region of chromosome 2 [Fátyol *et al.* (1994) *Nucl. Acids Res.* 22:3728-3736] was also used. Mouse chromosome

- 5 painting was done with a cloned long interspersed repetitive sequence [pMCP1.51] specific for the mouse euchromatin.

For cotransfection and for *in situ* hybridization, the pCH110 β -galactosidase construct [Pharmacia or Invitrogen], and λ cl 875 Sam7 phage DNA [New England Biolabs] were used.

10 2. Construction of Plasmid pPuroTel

Plasmid pPuroTel, which carries a Puromycin-resistance gene and a cloned 2.5 kb human telomeric sequence [see SEQ ID No. 3], was constructed from the pBabe-puro retroviral vector [Morgenstern *et al.* (1990) *Nucl. Acids Res.* 18:3587-3596; provided by Dr. L. Székely

- 15 (Microbiology and Tumorbiology Center, Karolinska Institutet, Stockholm); see, also Tonghua *et al.* (1995) *Chin. Med. J.* (Beijing, Engl. Ed.) 108:653-659; Couto *et al.* (1994) *Infect. Immun.* 62:2375-2378; Dunckley *et al.* (1992) *FEBS Lett.* 296:128-34; French *et al.* (1995) *Anal. Biochem.* 228:354-355; Liu *et al.* (1995) *Blood* 85:1095-1103;
- 20 International PCT application Nos. WO 9520044; WO 9500178, and WO 9419456].

F. Deposited cell lines

- Cell lines KE1-2/4, EC3/7C5, TF1004G19C5, 19C5xHa4, G3D5 and H1D3 have been deposited in accord with the Budapest Treaty at the
- 25 European Collection of Animal Cell Culture (ECACC) under Accession Nos. 96040924, 96040925, 96040926, 96040927, 96040928 and 96040929, respectively. The cell lines were deposited on April 9, 1996,

- at the European Collection of Animal Cell Cultures (ECACC) Vaccine Research and Production Laboratory, Public Health Laboratory Service, Centre for Applied Microbiology and Research, Porton Down, Salisbury, Wiltshire SP4 0JG, United Kingdom. The deposits were made in the
- 5 name of Gyula Hadlaczký of H. 6723, SZEGED, SZAMOS U.1.A. IX. 36. HUNGARY, who has authorized reference to the deposited cell lines in this application.

EXAMPLE 2

Preparation of EC3/7, EC3/7C5 and related cell lines

- 10 The EC3/7 cell line is an LMTK⁻ mouse cell line that contains the neo-centromere. The EC3/7C5 cell line is a single-cell subclone of EC3/7 that contains the neo-minichromosome.
- A. EC3/7 Cell line**
- As described in U.S. Patent No. 5,288,625 [see, also Praznovszky
- 15 et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046 and Hadlaczký et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110] *de novo* centromere formation occurs in a transformed mouse LMTK⁻ fibroblast cell line [EC3/7] after cointegration of λ constructs [λ CM8 and λ gtWESneo] carrying human and bacterial DNA.
- 20 By cotransfection of a 14 kb human DNA fragment cloned in λ [λ CM8] and a dominant marker gene [λ gtWESneo], a selectable centromere linked to a dominant marker gene [neo-centromere] was formed in mouse LMTK⁻ cell line EC3/7 [Hadlaczký et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110, see Figure 1]. Integration of the
- 25 heterologous DNA [the λ DNA and marker gene-encoding DNA] occurred into the short arm of an acrocentric chromosome [chromosome 7 (see, Figure 1B)], where an amplification process resulted in the formation of the new centromere [neo-centromere (see Figure 1C)]. On the dicentric chromosome (Figure 1C), the newly formed centromere region contains
- 30 the heterologous DNA (human, λ , and bacterial) introduced into the cell and an active centromere.

Having two functionally active centromeres on the same chromosome causes regular breakages between the centromeres [see, Figure 1E]. The distance between the two centromeres on the dicentric chromosome is estimated to be ~10-15 Mb, and the breakage that separates the minichromosome occurred between the two centromeres. Such specific chromosome breakages result in the appearance [in approximately 10% of the cells] of a chromosome fragment that carries the neo-centromere [Figure 1F]. This chromosome fragment is principally composed of human, λ , plasmid, and neomycin-resistance gene DNA, but it also has some mouse chromosomal DNA. Cytological evidence suggests that during the stabilization of the MMCneo, there was an inverted duplication of the chromosome fragment bearing the neo-centromere. The size of minichromosomes in cell lines containing the MMCneo is approximately 20-30 Mb; this finding indicates a two-fold increase in size.

From the EC3/7 cell line, which contains the dicentric chromosome [Figure 1E], two sublines [EC3/7C5 and EC3/7C6] were selected by repeated single-cell cloning. In these cell lines, the neo-centromere was found exclusively on a small chromosome [neo-minichromosome], while the formerly dicentric chromosome carried detectable amounts of the exogenously-derived DNA sequences but not an active neo-centromere [Figure 1F and 1G].

The minichromosomes of cell lines EC3/7C5 and EC3/7C6 are similar. No differences are detected in their architectures at either the cytological or molecular level. The minichromosomes were indistinguishable by conventional restriction endonuclease mapping or by long-range mapping using pulsed field electrophoresis and Southern hybridization. The cytoskeleton of cells of the EC3/7C6 line showed an increased sensitivity to colchicine, so the EC3/7C5 line was used for further detailed analysis.

B. Preparation of the EC3/7C5 and EC3/7C6 cell lines

The EC3/7C5 cells, which contain the neo-minichromosome, were produced by subcloning the EC3/7 cell line in high concentrations of G418 [40-fold the lethal dose] for 350 generations. Two single

- 5 cell-derived stable cell lines [EC3/7C5 and EC3/7C6] were established. These cell lines carry the neo-centromere on minichromosomes and also contain the remaining fragment of the dicentric chromosome. Indirect immunofluorescence with anti-centromere antibodies and subsequent in situ hybridization experiments demonstrated that the minichromosomes
10 derived from the dicentric chromosome. In EC3/7C5 and EC3/7C6 cell lines (140 and 128 metaphases, respectively) no intact dicentric chromosomes were found, and minichromosomes were detected in 97.2% and 98.1% of the cells, respectively. The minichromosomes have been maintained for over 150 cell generations. They do contain the
15 remaining portion of the formerly dicentric chromosome.

- Multiple copies of telomeric DNA sequences were detected in the marker centromeric region of the remaining portion of the formerly dicentric chromosome by in situ hybridization. This indicates that mouse telomeric sequences were coamplified with the foreign DNA sequences.
20 These stable minichromosome-carrying cell lines provide direct evidence that the extra centromere is functioning and is capable of maintaining the minichromosomes [see, U.S. Patent No. 5,288,625].

- The chromosome breakage in the EC3/7 cells, which separates the neo-centromere from the mouse chromosome, occurred in the G-band
25 positive "foreign" DNA region. This is supported by the observation of traces of λ and human DNA sequences at the broken end of the formerly dicentric chromosome. Comparing the G-band pattern of the chromosome fragment carrying the neo-centromere with that of the stable neo-minichromosome, reveals that the neo-minichromosome is an inverted
30 duplicate of the chromosome fragment that bears the neo-centromere. This is also evidenced by the observation that although the neo-

minichromosome carries only one functional centromere, both ends of the minichromosome are heterochromatic, and mouse satellite DNA sequences were found in these heterochromatic regions by *in situ* hybridization.

- 5 These two cell lines, EC3/7C5 and EC3/7C6, thus carry a selectable mammalian minichromosome [MMCneo] with a centromere linked to a dominant marker gene [Hadlaczký *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110]. MMCneo is intended to be used as a vector for minichromosome-mediated gene transfer and has been used as
- 10 model of a minichromosome-based vector system.

- Long range mapping studies of the MMCneo indicated that human DNA and the neomycin-resistance gene constructs integrated into the mouse chromosome separately, followed by the amplification of the chromosome region that contains the exogenous DNA. The MMCneo
- 15 contains about 30-50 copies of the λ CM8 and λ gtWESneo DNA in the form of approximately 160 kb repeated blocks, which together cover at least a 3.5 Mb region. In addition to these, there are mouse telomeric sequences [Praznovszky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046] and any DNA of mouse origin necessary for the correct
- 20 higher-ordered structural organization of chromatids.

- Using a chromosome painting probe mCPE1.51 [mouse long interspersed repeated DNA], which recognizes exclusively euchromatic mouse DNA, detectable amounts of interspersed repeat sequences were found on the MMCneo by *in situ* hybridization. The neo-centromere is
- 25 associated with a small but detectable amount of satellite DNA. The chromosome breakage that separates the neo-centromere from the mouse chromosome occurs in the "foreign" DNA region. This is demonstrated by the presence of λ and human DNA at the broken end of the formerly dicentric chromosome. At both ends of the MMCneo, however, there are
- 30 traces of mouse major satellite DNA as evidenced by *in situ* hybridization. This observation suggests that the doubling in size of the chromosome

fragment carrying the neo-centromere during the stabilization of the MMCneo is a result of an inverted duplication. Although mouse telomere sequences, which coamplified with the exogenous DNA sequences during the neo-centromere formation, may provide sufficient telomeres for the

- 5 MMCneo, the duplication could have supplied the functional telomeres for the minichromosome.

The nucleotide sequence of portions of the neo-minichromosomes was determined as follows. Total DNA was isolated from EC3/7C5 cells according to standard procedures. The DNA was subjected to nucleic

- 10 acid amplification using the Expand Long Template PCR system [Boehringer Mannheim] according to the manufacturer's procedures. The amplification procedure required only a single 33-mer oligonucleotide primer corresponding to sequence in a region of the phage λ right arm, which is contained in the neo-minichromosome. The sequence of this

- 15 oligonucleotide is set forth as the first 33 nucleotides of SEQ ID No. 13. Because the neo-minichromosome contains a series of inverted repeats of this sequence, the single oligonucleotide was used as a forward and reverse primer resulting in amplification of DNA positioned between sets of inverted repeats of the phage λ DNA. Three products were obtained
20 from the single amplification reaction, which suggests that the sequence of the DNA located between different sets of inverted repeats may differ. In a repeating nucleic acid unit within an artificial chromosome, minor differences may be present and may occur during culturing of cells containing the artificial chromosome. For example, base pair changes
25 may occur as well as integration of mobile genetic elements and deletions of repeated sequences.

- Each of the three products was subjected to DNA sequence analysis. The sequences of the three products are set forth in SEQ ID Nos. 13, 14, and 15, respectively. To be certain that the sequenced
30 products were amplified from the neo-minichromosome, control amplifications were conducted using the same primers on DNA isolated

from negative control cell lines (mouse Ltk⁺ cells) lacking minichromosomes and the formerly dicentric chromosome, and positive control cell lines [the mouse-hamster hybrid cell line GB43 generated by treating 19C5xHa4 cells (see Figure 4) with BrdU followed by growth in

5 G418-containing selective medium and retreatment with BrdU] containing the neo-minichromosome only. Only the positive control cell line yielded the three amplification products; no amplification product was detected in the negative control reaction. The results obtained in the positive control amplification also demonstrate that the neo-minichromosome DNA, and

10 not the fragment of the formerly dicentric mouse chromosome, was amplified.

The sequences of the three amplification products were compared to those contained in the Genbank/EMBL database. SEQ ID Nos. 13 and 14 showed high (~96%) homology to portions of DNA from intracisternal

15 A-particles from mouse. SEQ ID No. 15 showed no significant homology with sequences available in the database. All three of these sequences may be used for generating gene targeting vectors as homologous DNAs to the neo-minichromosome.

C. Isolation and partial purification of minichromosomes

20 Mitotic chromosomes of EC3/7C5 cells were isolated as described by Hadlaczký et al. [(1981) Chromosoma 81:537-555], using a glycine-hexylene glycol buffer system [Hadlaczký et al. (1982) Chromosoma 86:643-659]. Chromosome suspensions were centrifuged at 1,200 x g for 30 minutes. The supernatant containing

25 minichromosomes was centrifuged at 5,000 x g for 30 minutes and the pellet was resuspended in the appropriate buffer. Partially purified minichromosomes were stored in 50% glycerol at -20° C.

D. Stability of the MMCneo maintenance and *neo* expression

30 EC3/7C5 cells grown in non-selective medium for 284 days and then transferred to selective medium containing 400 µg/ml G418 showed a 96% plating efficiency (colony formation) compared to control cells

cultured permanently in the presence of G418. Cytogenetic analysis indicated that the MMCneo is stably maintained at one copy per cell under selective and non-selective culture conditions. Only two metaphases with two MMCneo were found in 2,270 metaphases

5 analyzed.

Southern hybridization analysis showed no detectable changes in DNA restriction patterns, and similar hybridization intensities were observed with a *neo* probe when DNA from cells grown under selective or non-selective culture conditions were compared.

- 10 Northern analysis of RNA transcripts from the *neo* gene isolated from cells grown under selective and non-selective conditions showed only minor and not significant differences. Expression of the *neo* gene persisted in EC3/7C5 cells maintained in F-12 medium free of G418 for 290 days under non-selective culture conditions. The long-term
- 15 expression of the *neo* gene(s) from the minichromosome may be influenced by the nuclear location of the MMCneo. *In situ* hybridization experiments revealed a preferential peripheral location of the MMCneo in the interphase nucleus. In more than 60% of the 2,500 nuclei analyses, the minichromosome was observed at the perimeter of the nucleus near
- 20 the nuclear envelope.

EXAMPLE 3

Minichromosome transfer and production of the λ -neo-chromosome

A. Minichromosome transfer

- The neo-minichromosome [referred to as MMCneo, FIG. 2C] has
- 25 been used for gene transfer by fusion of minichromosome-containing cells [EC3/7C5 or EC3/7C6] with different mammalian cells, including hamster and human. Thirty-seven stable hybrid cell lines have been produced. All established hybrid cell lines proved to be true hybrids as evidenced by *in situ* hybridization using biotinylated human, and hamster genomic, or
- 30 pMCPE1.51 mouse long interspersed repeated DNA probes for "chromosome painting". The MMCneo has also been successfully

transferred into mouse A9, L929 and pluripotent F9 teratocarcinoma cells by fusion of microcells derived from EC3/7C5 cells. Transfer was confirmed by PCR, Southern blotting and *in situ* hybridization with minichromosome-specific probes. The cytogenetic analysis confirmed

- 5 that, as expected for microcell fusion, a few cells [1-5%] received [or retained] the MMCneo.

These results demonstrate that the MMCneo is tolerated by a wide range of cells. The prokaryotic genes and the extra dosage for the human and λ sequences carried on the minichromosome seem to be not

- 10 disadvantageous for tissue culture cells.

The MMCneo is the smallest chromosome of the EC3/7C5 genome and is estimated to be approximately 20-30 Mb, which is significantly smaller than the majority of the host cell (mouse) chromosomes. By virtue of the smaller size, minichromosomes can be partially purified from

- 15 a suspension of isolated chromosomes by a simple differential centrifugation. In this way, minichromosome suspensions of 15-20% purity have been prepared. These enriched minichromosome preparations can be used to introduce, such as by microinjection or lipofection, the minichromosome into selected target cells. Target cells include

- 20 therapeutic cells that can be use in methods of gene therapy, and also embryonic cells for the preparation of transgenic (non-human) animals.

The MMCneo is capable of autonomous replication, is stably maintained in cells, and permits persistent expression of the *neo* gene(s), even after long-term culturing under non-selective conditions. It is a

25 non-integrative vector that appears to occupy a territory near the nuclear envelope. Its peripheral localization in the nucleus may have an important role in maintaining the functional integrity and stability of the MMCneo. Functional compartmentalization of the host nucleus may have an effect on the function of foreign sequences. In addition, MMCneo contains

30 megabases of λ DNA sequences that should serve as a target site for homologous recombination and thus integration of desired gene(s) into

the MMCneo. It can be transferred by cell and microcell fusion, microinjection, electroporation, lipid-mediated carrier systems or chromosome uptake. The neo-centromere of the MMCneo is capable of maintaining and supporting the normal segregation of a larger 150-200 Mb λ neo-chromosome. This result demonstrates that the MMCneo chromosome should be useful for carrying large fragments of heterologous DNA.

B. Production of the λ neo-chromosome

- In the hybrid cell line KE1-2/4 made by fusion of EC3/7 and Chinese hamster ovary cells [FIG 2], the separation of the neo-centromere from the dicentric chromosome was associated with a further amplification process. This amplification resulted in the formation of a stable chromosome of average size [i.e., the λ neo-chromosome; see, Praznovsky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046]. The λ neo-chromosome carries a terminally located functional centromere and is composed of seven large amplicons containing multiple copies of λ , human, bacterial, and mouse DNA sequences [see FIG 2]. The amplicons are separated by mouse major satellite DNA [Praznovsky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:11042-11046] which forms narrow bands of constitutive heterochromatin between the amplicons.

EXAMPLE 4

Formation of the "sausage chromosome" [SC]

- The findings set forth in the above EXAMPLES demonstrate that the centromeric region of the mouse chromosome 7 has the capacity for large-scale amplification [other results indicate that this capacity is not unique to chromosome 7]. This conclusion is further supported by results from cotransfection experiments, in which a second dominant selectable marker gene and a non-selected marker gene were introduced into EC3/7C5 cells carrying the formerly dicentric chromosome 7 and the neo-minichromosome. The EC3/7C5 cell line was transformed with λ phage DNA, a hygromycin-resistance gene construct [pH132], and a β -

galactosidase gene construct [pCH110]. Stable transformants were selected in the presence of high concentrations [400 μ g/ml] Hygromycin B, and analyzed by Southern hybridization. Established transformant cell lines showing multiple copies of integrated exogenous DNA were studied by *in situ* hybridization to localize the integration site(s), and by LacZ staining to detect β -galactosidase expression.

A. Materials and methods

1. Construction of pH132

The pH132 plasmid carries the hygromycin B resistance gene and the anti-HIV-1 *gag* ribozyme [see, SEQ ID NO. 6 for DNA sequence that corresponds to the sequence of the ribozyme] under control of the β -actin promoter. This plasmid was constructed from pHyg plasmid [Sugden *et al.* (1985) *Mol. Cell. Biol.* 5:410-413; a gift from Dr. A. D. Riggs, Beckman Research Institute, Duarte; see, also, *e.g.*, U.S. Patent No. 4,997,764], and from pPC-RAG12 plasmid [see, Chang *et al.* (1990) *Clin. Biotech.* 2:23-31; provided by Dr. J. J. Rossi, Beckman Research Institute, Duarte; see, also U.S. Patent Nos. 5,272,262, 5,149,796 and 5,144,019, which describes the anti-HIV *gag* ribozyme and construction of a mammalian expression vector containing the ribozyme insert linked to the β -actin promoter and SV40 late gene transcriptional termination and polyA signals]. Construction of pPC-RAG12 involved insertion of the ribozyme insert flanked by BamHI linkers was into BamHI-digested pH β -Apr-1gpt [see, Gunning *et al.* (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84:4831-4835, see, also U.S. Patent No. 5,144,019].

Plasmid pH132 was constructed as follows. First, pPC-RAG12 [described by Chang *et al.* (1990) *Clin. Biotech.* 2:23-31] was digested with BamHI to excise a fragment containing an anti-HIV ribozyme gene [referred to as ribozyme D by Chang *et al.* [(1990) *Clin. Biotech.* 2:23-31]; see also U.S. Patent No. 5,144,019 to Rossi *et al.*, particularly Figure 4 of the patent] flanked by the human β -actin promoter at the 5' end of the gene and the SV40 late transcriptional termination and

polyadenylation signals at the 3' end of the gene. As described by Chang et al. [(1990) Clin. Biotech. 2:23-31], ribozyme D is targeted for cleavage of the translational initiation region of the HIV gag gene. This fragment of pPC-RAG12 was subcloned into pBluescript-KS(+) [Stratagene, La Jolla, CA] to produce plasmid 132. Plasmid 132 was then digested with XhoI and EcoRI to yield a fragment containing the ribozyme D gene flanked by the β -actin promoter at the 5' end and the SV40 termination and polyadenylation signals at the 3' end of the gene. This fragment was ligated to the largest fragment generated by digestion of pHyg [Sugden et al. (1985) Mol. Cell. Biol. 5:410-413] with EcoRI and Sall to yield pH132. Thus, pH132 is an ~9.3 kb plasmid containing the following elements: the β -actin promoter linked to an anti-HIV ribozyme gene followed by the SV40 termination and polyadenylation signals, the thymidine kinase gene promoter linked to the hygromycin-resistance gene followed by the thymidine kinase gene polyadenylation signal, and the E. coli ColE1 origin of replication and the ampicillin-resistance gene.

The plasmid pHyg [see, e.g., U.S. Patent Nos. 4,997,764, 4,686,186 and 5,162,215], which confers resistance to hygromycin B using transcriptional controls from the HSV-1 tk gene, was originally constructed from pKan2 [Yates et al. (1984) Proc. Natl. Acad. Sci. U.S.A. 81:3806-3810] and pLG89 [see, Gritz et al. (1983) Gene 25:179-188]. Briefly pKan2 was digested with SmaI and BglII to remove the sequences derived from transposon Tn5. The hygromycin-resistance hph gene was inserted into the digested pKan2 using blunt-end ligation at the SnaI site and "sticky-end" ligation [using 1 Weiss unit of T4 DNA ligase (BRL) in 20 microliter volume] at the BglII site. The SmaI and BglII sites of pKan2 were lost during ligation.

The resulting plasmid pH132, produced from introduction of the anti-HIV ribozyme construct with promoter and polyA site into pHyg, includes the anti-HIV ribozyme under control of the β -actin promoter as well as the hygromycin-resistance gene under control of the TK promoter.

2. Chromosome banding

Trypsin G-banding of chromosomes was performed as described in EXAMPLE 1.

3. Cell cultures

- 5 TF1004G19 and TF1004G-19C5 mouse cells and the 19C5xHa4 hybrid, described below, and its sublines were cultured in F-12 medium containing 400 $\mu\text{g/ml}$ Hygromycin B [Calbiochem].

B. Cotransfection of EC3/7C5 to produce TF1004G19

- Cotransfection of EC3/7C5 cells with plasmids [pH132, pCH110
10 available from Pharmacia, see, also Hall et al. (1983) J. Mol. Appl. Gen. 2:101-109] and with λ DNA [λ cl 875 Sam 7(New England Biolabs)] was conducted using the calcium phosphate DNA precipitation method [see, e.g., Chen et al. (1987) Mol. Cell. Biol. 7:2745-2752], using 2-5 μg plasmid DNA and 20 μg λ phage DNA per 5×10^6 recipient cells.

15 C. Cell lines containing the sausage chromosome

- Analysis of one of the transformants, designated TF1004G19, revealed that it has a high copy number of integrated pH132 and pCH110 sequences, and a high level of β -galactosidase expression. G-banding and *in situ* hybridization with a human probe [CM8; see, e.g., U.S. application
20 Serial No. 08/375,271] revealed unexpectedly that integration had occurred in the formerly dicentric chromosome 7 of the EC3/7C5 cell line. Furthermore, this chromosome carried a newly formed heterochromatic chromosome arm. The size of this heterochromatic arm varied between ~150 and ~800 Mb in individual metaphases.

- 25 By single cell cloning from the TF1004G19 cell line, a subclone TF1004G-19C5 [FIG 2D], which carries a stable chromosome 7 with a ~100-150 Mb heterochromatic arm [the sausage chromosome] was obtained. This cell line has been deposited in the ECACC under Accession No. 96040926. This chromosome arm is composed of four to
30 five satellite segments rich in satellite DNA, and evenly spaced integrated heterologous "foreign" DNA sequences. At the end of the compact

heterochromatic arm of the sausage chromosome, a less condensed euchromatic terminal segment is regularly observed. This subclone was used for further analyses.

D. Demonstration that the sausage chromosome is derived from the formerly dicentric chromosome

- 5 *In situ* hybridization with λ phage and pH132 DNA on the TF1004G-19C5 cell line showed positive hybridization only on the minichromosome and on the heterochromatic arm of the "sausage" chromosome [Fig. 2D]. It appears that the "sausage" chromosome
- 10 [herein also referred to as the SC] developed from the formerly dicentric chromosome (FD) of the EC3/7C5 cell line.

- To establish this, the integration sites of pCH110 and pH132 plasmids were determined. This was accomplished by *in situ* hybridization on these cells with biotin-labeled subfragments of the
- 15 hygromycin-resistance gene and the β -galactosidase gene. Both experiments resulted in narrow hybridizing bands on the heterochromatic arm of the sausage chromosome. The same hybridization pattern was detected on the sausage chromosome using a mixture of biotin-labeled λ probe and pH132 plasmid, proving the cointegration of λ phages, pH132
- 20 and pCH110 plasmids.

- To examine this further, the cells were cultured in the presence of the DNA-binding dye Hoechst 33258. Culturing of mouse cells in the presence of this dye results in under-condensation of the pericentric heterochromatin of metaphase chromosomes, thereby permitting better
- 25 observation of the hybridization pattern. Using this technique, the heterochromatic arm of the sausage chromosome of TF1004G-19C5 cells showed regular under-condensation revealing the details of the structure of the "sausage" chromosome by *in situ* hybridization. Results of *in situ* hybridization on Hoechst-treated TF1004G-19C5 cells with biotin-labeled
- 30 subfragments of hygromycin-resistance and β -galactosidase genes shows that these genes are localized only in the heterochromatic arm of the

sausage chromosome. In addition, an equal banding hybridization pattern was observed. This pattern of repeating units [amplicons] clearly indicates that the sausage chromosome was formed by an amplification process and that the λ phage, pH132 and pCH110 plasmid DNA

5 sequences border the amplicons.

In another series of experiments using fluorescence *in situ* hybridization [FISH] carried out with mouse major satellite DNA, the main component of the mouse pericentric heterochromatin, the results confirmed that the amplicons of the sausage chromosome are primarily
10 composed of satellite DNA.

E. The sausage chromosome has one centromere

To determine whether mouse centromeric sequences had participated in the amplification process forming the "sausage" chromosome and whether or not the amplicons carry inactive
15 centromeres, *in situ* hybridization was carried out with mouse minor satellite DNA. Mouse minor satellite DNA is localized specifically near the centromeres of all mouse chromosomes. Positive hybridization was detected in all mouse centromeres including the sausage chromosome, which, however, only showed a positive signal at the beginning of the
20 heterochromatic arm.

Indirect immunofluorescence with a human anti-centromere antibody [LU 851] which recognizes only functional centromeres [see, e.g., Hadlaczy et al. (1989) Chromosoma 97:282-288] proved that the sausage chromosome has only one active centromere. The centromere
25 comes from the formerly dicentric part of the chromosome and co-localizes with the *in situ* hybridization signal of the mouse minor DNA probe.

F. The selected and non-selected heterologous DNA in the heterochromatin of the sausage chromosome is expressed

1. High levels of the heterologous genes are expressed

30 The TF1004G-19C5 cell line thus carries multiple copies of hygromycin-resistance and β -galactosidase genes localized only in the

heterochromatic arm of the sausage chromosome. The TF1004G-19C5 cells can grow very well in the presence of 200 $\mu\text{g/ml}$ or even 400 $\mu\text{g/ml}$ hygromycin B. [The level of expression was determined by Northern hybridization with a subfragment of the hygromycin-resistance gene and single copy gene.]

The expression of the non-selected β -galactosidase gene in the TF1004G-19C5 transformant was detected with LacZ staining of the cells. By this method one hundred percent of the cells stained dark blue, showing that there is a high level of β -galactosidase expression in all of TF1004G-19C5 cells.

2. The heterologous genes that are expressed are in the heterochromatin of the sausage chromosome

To demonstrate that the genes localized in the constitutive heterochromatin of the sausage chromosome provide the hygromycin resistance and the LacZ staining capability of TF1004G-19C5 transformants [i.e. β -gal expression], PEG-induced cell fusion between TF1004G-19C5 mouse cells and Chinese hamster ovary cells was performed. The hybrids were selected and maintained in HAT medium containing G418 [400 $\mu\text{g/ml}$] and hygromycin [200 $\mu\text{g/ml}$]. Two hybrid clones designated 19C5xHa3 and 19C5xHa4, which have been deposited in the ECACC under Accession No. 96040927, were selected. Both carry the sausage chromosome and the minichromosome.

Twenty-seven single cell derived colonies of the 19C5xHa4 hybrid were maintained and analyzed as individual subclones. *In situ* hybridization with hamster and mouse chromosome painting probes and hamster chromosome 2-specific probes verified that the 19C5xHa4 clone contains the complete Chinese hamster genome and a partial mouse genome. All 19C5xHa4 subclones retained the hamster genome, but different subclones showed different numbers of mouse chromosomes indicating the preferential elimination of mouse chromosomes.

To promote further elimination of mouse chromosomes, hybrid cells were repeatedly treated with BrdU. The BrdU treatments, which destabilize the genome, result in significant loss of mouse chromosomes. The BrdU-treated 19C5xHa4 hybrid cells were divided to three groups.

- 5 One group of the hybrid cells (GH) were maintained in the presence of hygromycin (200 $\mu\text{g/ml}$) and G418 (400 $\mu\text{g/ml}$), and the other two groups of the cells were cultured under G418 (G) or hygromycin (H) selection conditions to promote the elimination of the sausage chromosome or minichromosome.
- 10 One month later, single cell derived subclones were established from these three subcultures of the 19C5xHa4 hybrid line. The subclones were monitored by *in situ* hybridization with biotin-labeled λ phage and hamster chromosome painting probes. Four individual clones [G2B5, G3C5, G4D6, G2B4] selected in the presence of G418 that had lost the
- 15 sausage chromosome but retained the minichromosome were found. Under hygromycin selection only one subclone [H1D3] lost the minichromosome. In this clone the megachromosome [see Example 5] was present.

- Since hygromycin-resistance and β -galactosidase genes were
- 20 thought to be expressed from the sausage chromosome, the expression of these genes was analyzed in the four subclones that had lost the sausage chromosome. In the presence of 200 $\mu\text{g/ml}$ hygromycin, one hundred percent of the cells of four individual subclones died. In order to detect the β -galactosidase expression hybrid, subclones were analyzed by
- 25 LacZ staining. One hundred percent of the cells of the four subclones that lost the sausage chromosome also lost the LacZ staining capability. All of the other hybrid subclones that had not lost the sausage chromosome under the non-selective culture conditions showed positive LacZ staining.

- 30 These findings demonstrate that the expression of hygromycin-resistance and β -galactosidase genes is linked to the presence of the

sausage chromosome. Results of *in situ* hybridizations show that the heterologous DNA is expressed from the constitutive heterochromatin of the sausage chromosome.

- In situ* hybridization studies of three other hybrid subclones [G2C6, 5 G2D1, and G4D5] did not detect the presence of the sausage chromosome. By the LacZ staining method, some stained cells were detected in these hybrid lines, and when these subclones were transferred to hygromycin selection some colonies survived. Cytological analysis and *in situ* hybridization of these hygromycin-resistant colonies 10 revealed the presence of the sausage chromosome, suggesting that only the cells of G2C6, G2D1 and G4D5 hybrids that had not lost the sausage chromosome were able to preserve the hygromycin resistance and β -galactosidase expression. These results confirmed that the expression of these genes is linked to the presence of the sausage chromosome. The 15 level of β -galactosidase expression was determined by the immunoblot technique using a monoclonal antibody.

- Hygromycin resistance and β -galactosidase expression of the cells which contained the sausage chromosome were provided by the genes localized in the mouse pericentric heterochromatin. This was 20 demonstrated by performing Southern DNA hybridizations on the hybrid cells that lack the sausage chromosome using PCR-amplified subfragments of hygromycin-resistance and β -galactosidase genes as probes. None of the subclones showed hybridization with these probes; however, all of the analyzed clones contained the minichromosome. 25 Other hybrid clones that contain the sausage chromosome showed intense hybridization with these DNA probes. These results lead to the conclusion that hygromycin resistance and β -galactosidase expression of the cells that contain the sausage chromosome were provided by the genes localized in the mouse pericentric heterochromatin.

EXAMPLE 5

The gigachromosome

As described in Example 4, the sausage chromosome was transferred into Chinese hamster cells by cell fusion. Using Hygromycin B/HAT and G418 selection, two hybrid clones 19C5xHa3 and 19C5xHa4 were produced that carry the sausage chromosome. *In situ* hybridization, using hamster and mouse chromosome-painting probes and a hamster chromosome 2-specific probe, verified that clone 19C5xHa4 contains a complete Chinese hamster genome as well as partial mouse genomes.

- 10 Twenty-seven separate colonies of 19C5xHa4 cells were maintained and analyzed as individual subclones. Twenty-six out of 27 subclones contained a morphologically unchanged sausage chromosome.

- In one subclone of the 19C5xHa3 cell line, 19C5xHa47 [see Fig. 2E], the heterochromatic arm of the sausage chromosome became
15 unstable and showed continuous intrachromosomal growth. In extreme cases, the amplified chromosome arm exceeded 1000 Mb in size (gigachromosome).

EXAMPLE 6

The stable megachromosome

- 20 A. Generation of cell lines containing the megachromosome

- All 19C5xHa4 subclones retained a complete hamster genome, but different subclones showed different numbers of mouse chromosomes, indicating the preferential elimination of mouse chromosomes. As described in Example 4, to promote further elimination of mouse
25 chromosomes, hybrid cells were treated with BrdU, cultured under G418 (G) or hygromycin (H) selection conditions followed by repeated treatment with 10^{-4} M BrdU for 16 hours and single cell subclones were established. The BrdU treatments appeared to destabilize the genome, resulting in a change in the sausage chromosome as well. A gradual increase in a cell
30 population in which a further amplification had occurred was observed. In addition to the ~100-150 Mb heterochromatic arm of the sausage

chromosome, an extra centromere and a ~ 150-250 Mb heterochromatic chromosome arm were formed, which differed from those of mouse chromosome 7. By the acquisition of another euchromatic terminal segment, a new submetacentric chromosome (megachromosome) was formed. Seventy-nine individual subclones were established from these BrdU-treated cultures by single-cell cloning: 42 subclones carried the intact megachromosome, 5 subclones carried the sausage chromosome, and in 32 subclones fragments or translocated segments of the megachromosome were observed. Twenty-six subclones that carried the megachromosome were cultured under non-selective conditions over a two-month period. In 19 out of 26 subclones, the megachromosome was retained. Those subclones which lost the megachromosomes all became sensitive to Hygromycin B and had no β -galactosidase expression, indicating that both markers were linked to the megachromosome.

Two sublines (G3D5 and H1D3), which were chosen for further experiments, showed no changes in the morphology of the megachromosome during more than 100 generations under selective conditions. The G3D5 cells had been obtained by growth of 19C5xHa4 cells in G418-containing medium followed by repeated BrdU treatment, whereas H1D3 cells had been obtained by culturing 19C5xHa4 cells in hygromycin-containing medium followed by repeated BrdU treatment.

B. Structure of the megachromosome

The following results demonstrate that, apart from the euchromatic terminal segments, the integrated foreign DNA (and as in the exemplified embodiments, rDNA sequence), the whole megachromosome is constitutive heterochromatin, containing a tandem array of at least 40 [~ 7.5 Mb] blocks of mouse major satellite DNA [see Figures 2 and 3]. Four satellite DNA blocks are organized into a giant palindrome [amplicon] carrying integrated exogenous DNA sequences at each end. The long and short arms of the submetacentric megachromosome contains 6 and 4 amplicons, respectively. It is of course understood that the specific

organization and size of each component can vary among species, and also the chromosome in which the amplification event initiates.

1. The megachromosome is composed primarily of heterochromatin

- 5 Except for the terminal regions and the integrated foreign DNA, the megachromosome is composed primarily of heterochromatin. This was demonstrated by C-banding of the megachromosome, which resulted in positive staining characteristic of constitutive heterochromatin. Apart from the terminal regions and the integrated foreign DNA, the whole
- 10 megachromosome appears to be heterochromatic. Mouse major satellite DNA is the main component of the pericentric, constitutive heterochromatin of mouse chromosomes and represents ~10% of the total DNA [Waring et al. (1966) Science 154:791-794]. Using a mouse major satellite DNA probe for *in situ* hybridization, strong hybridization
- 15 was observed throughout the megachromosome, except for its terminal regions. The hybridization showed a segmented pattern: four large blocks appeared on the short arm and usually 4-7 blocks were seen on the long arm. By comparing these segments with the pericentric regions of normal mouse chromosomes that carry ~15 Mb of major satellite DNA, the size
- 20 of the blocks of major satellite DNA on the megachromosome was estimated to be ~30 Mb.

- Using a mouse probe specific to euchromatin [pMCPE1.51; a mouse long interspersed repeated DNA probe], positive hybridization was detected only on the terminal segments of the megachromosome of the
- 25 H1D3 hybrid subline. In the G3D5 hybrids, hybridization with a hamster-specific probe revealed that several megachromosomes contained terminal segments of hamster origin on the long arm. This observation indicated that the acquisition of the terminal segments on these chromosomes happened in the hybrid cells, and that the long arm of the
- 30 megachromosome was the recently formed one arm. When a mouse minor satellite probe was used, specific to the centromeres of mouse

chromosomes [Wong *et al.* (1988) *Nucl. Acids Res.* 16:11645-11661], a strong hybridization signal was detected only at the primary constriction of the megachromosome, which colocalized with the positive immunofluorescence signal produced with human anti-centromere serum [LU851].

- 5 *In situ* hybridization experiments with pH132, pCH110, and λ DNA probes revealed that all heterologous DNA was located in the gaps between the mouse major satellite DNA segments. Each segment of mouse major satellite DNA was bordered by a narrow band of integrated heterologous DNA, except at the second segment of the long arm where
- 10 a double band of heterologous DNA existed, indicating that the major satellite DNA segment was missing or considerably reduced in size here. This chromosome region served as a useful cytological marker in identifying the long arm of the megachromosome. At a frequency of 10^{-4} , "restoration" of these missing satellite DNA blocks was observed in
- 15 one chromatid, when the formation of a whole segment on one chromatid occurred.

- After Hoechst 33258 treatment (50 μ g/ml for 16 hours), the megachromosome showed undercondensation throughout its length except for the terminal segments. This made it possible to study the
- 20 architecture of the megachromosome at higher resolution. *In situ* hybridization with the mouse major satellite probe on undercondensed megachromosomes demonstrated that the ~ 30 Mb major satellite segments were composed of four blocks of ~ 7.5 Mb separated from each other by a narrow band of non-hybridizing sequences [Figure 3].
- 25 Similar segmentation can be observed in the large block of pericentric heterochromatin in metacentric mouse chromosomes from the LMTK⁻ and A9 cell lines.

2. **The megachromosome is composed of segments containing two tandem ~ 7.5 Mb blocks followed by two inverted blocks**
- 30

Because of the asymmetry in thymidine content between the two strands of the DNA of the mouse major satellite, when mouse cells are

grown in the presence of BrdU for a single S phase, the constitutive heterochromatin shows lateral asymmetry after FPG staining. Also, in the 19C5xHa4 hybrids, the thymidine-kinase [*Tk*] deficiency of the mouse fibroblast cells was complemented by the hamster *Tk* gene, permitting

5 BrdU incorporation experiments.

A striking structural regularity in the megachromosome was detected using the FPG technique. In both chromatids, alternating dark and light staining that produced a checkered appearance of the megachromosome was observed. A similar picture was obtained by

- 10 labelling with fluorescein-conjugated anti-BrdU antibody. Comparing these pictures to the segmented appearance of the megachromosome showed that one dark and one light FPG band corresponded to one ~30 Mb segment of the megachromosome. These results suggest that the two halves of the ~30 Mb segment have an inverted orientation. This
- 15 was verified by combining *in situ* hybridization and immunolabelling of the incorporated BrdU with fluorescein-conjugated anti-BrdU antibody on the same chromosome. Since the ~30 Mb segments [or amplicons] of the megachromosome are composed of four blocks of mouse major satellite DNA, it can be concluded that two tandem ~7.5 Mb blocks are followed
- 20 by two inverted blocks within one segment.

Large-scale mapping of megachromosome DNA by pulsed-field electrophoresis and Southern hybridization with "foreign" DNA probes revealed a simple pattern of restriction fragments. Using endonucleases with none, or only a single cleavage site in the integrated foreign DNA

25 sequences, followed by hybridization with a *hyg* probe, 1-4 predominant fragments were detected. Since the megachromosome contains 10-12 amplicons with an estimated 3-8 copies of *hyg* sequences per amplicon (30-90 copies per megachromosome), the small number of hybridizing fragments indicates the homogeneity of DNA in the amplified segments.

3. Scanning electron microscopy of the megachromosome confirmed the above findings

The homogeneous architecture of the heterochromatic arms of the megachromosome was confirmed by high resolution scanning electron
 5 microscopy. Extended arms of megachromosomes, and the pericentric heterochromatic region of mouse chromosomes, treated with Hoechst 33258, showed similar structure. The constitutive heterochromatic regions appeared more compact than the euchromatic segments. Apart from the terminal regions, both arms of the megachromosome were
 10 completely extended, and showed faint grooves, which should correspond to the border of the satellite DNA blocks in the non-amplified chromosomes and in the megachromosome. Without Hoechst treatment, the grooves seemed to correspond to the amplicon borders on the megachromosome arms. In addition, centromeres showed a more
 15 compact, finely fibrous appearance than the surrounding heterochromatin.

4. The megachromosome of 1B3 cells contains rRNA gene sequence

The sequence of the megachromosome in the region of the sites of integration of the heterologous DNA was investigated by isolation of
 20 these regions through using cloning methods and sequence analysis of the resulting clones. The results of this analysis revealed that the heterologous DNA was located near mouse ribosomal RNA gene (i.e., rDNA) sequences contained in the megachromosome.

25 a. Cloning of regions of the megachromosomes in which heterologous DNA had integrated

Megachromosomes were isolated from 1B3 cells (which were generated by repeated BrdU treatment and single cell cloning of H1xHE41 cells (see Figure 4) and which contain a truncated megachromosome) using fluorescence-activated cell sorting methods as described herein (see
 30 Example 10). Following separation of the SATACs (megachromosomes) from the endogenous chromosomes, the isolated megachromosomes were stored in GH buffer (100 mM glycine, 1% hexylene glycol, pH 8.4-8.6

adjusted with saturated calcium hydroxide solution; see Example 10) and centrifuged into an agarose bed in 0.5 M EDTA.

Large-scale mapping of the megachromosome around the area of the site of integration of the heterologous DNA revealed that it is enriched in sequence containing rare-cutting enzyme sites, such as the recognition site for NotI. Additionally, mouse major satellite DNA (which makes up the majority of the megachromosome) does not contain NotI recognition sites. Therefore, to facilitate isolation of regions of the megachromosome associated with the site of integration of the heterologous DNA, the isolated megachromosomes were cleaved with NotI, a rare cutting restriction endonuclease with an 8-bp GC recognition site. Fragments of the megachromosome were inserted into plasmid pWE15 (Stratagene, La Jolla, California) as follows. Half of a 100- μ l low melting point agarose block (mega-plug) containing the isolated SATACs was digested with NotI overnight at 37°C. Plasmid pWE15 was similarly digested with NotI overnight. The mega-plug was then melted and mixed with the digested plasmid, ligation buffer and T4 ligase. Ligation was conducted at 16°C overnight. Bacterial DH5 α cells were transformed with the ligation product and transformed cells were plated onto LB/Amp plates. Fifteen to twenty colonies were grown on each plate for a total of 189 colonies. Plasmid DNA was isolated from colonies that survived growth on LB/Amp medium and was analyzed by Southern blot hybridization for the presence of DNA that hybridized to a pUC19 probe. This screening methodology assured that all clones, even clones lacking an insert but yet containing the pWE15 plasmid, would be detected. Any clones containing insert DNA would be expected to contain non-satellite, GC-rich megachromosome DNA sequences located at the site of integration of the heterologous DNA. All colonies were positive for hybridizing DNA.

Liquid cultures of all 189 transformants were used to generate cosmid minipreps for analysis of restriction sites within the insert DNA. Six of the original 189 cosmid clones contained an insert. These clones

were designated as follows: 28 (~9-kb insert), 30 (~9-kb insert), 60 (~4-kb insert), 113 (~9-kb insert), 157 (~9-kb insert) and 161 (~9-kb insert). Restriction enzyme analysis indicated that three of the clones (113, 157 and 161) contained the same insert.

5 **b. *In situ* hybridization experiments using isolated segments of the megachromosome as probes**

- Insert DNA from clones 30, 113, 157 and 161 was purified, labeled and used as probes in *in situ* hybridization studies of several cell lines. Counterstaining of the cells with propidium iodide facilitated
- 10 identification of the cytological sites of the hybridization signals. The locations of the signals detected within the cells are summarized in the following table:

CELL TYPE	PROBE	LOCATION OF SIGNAL
Human Lymphocyte (male)	No. 161	4-5 pairs of acrocentric chromosomes at centromeric regions.
Mouse Spleen	No. 161	Acrocentric ends of 4 pairs of chromosomes.
EC3/7C5 Cells	No. 161	Minichromosome and the end of the formerly dicentric chromosome. Pericentric heterochromatin of one of the metacentric mouse chromosomes. Centromeric region of some of the other mouse chromosomes.
K20 Chinese Hamster Cells	No. 30	Ends of at least 6 pairs of chromosomes. An interstitial signal on a short chromosome.
HB31 Cells (mouse-hamster hybrid cells derived from H1D3 cells by repeated BrdU treatment and single cell cloning which carries the megachromosome)	No. 30	Acrocentric ends of at least 12 pairs of chromosomes. Centromeres of certain chromosomes and the megachromosome. Borders of the amplicons of the megachromosome.
Mouse Spleen Cells	No. 30	Similar to signal observed for probe no. 161. Centromeres of 5 pairs of chromosomes. Weak cross-hybridization to pericentric heterochromatin.

CELL TYPE	PROBE	LOCATION OF SIGNAL
HB31 Cells	No. 113	Similar to signal observed for probe no. 30.
Mouse Spleen Cells	No. 113	Centromeric region of 5 pairs of chromosomes.
K20 Cells	No. 113	At least 6 pairs of chromosomes. Weak signal at some telomeres and several interspersed signals.
Human Lymphocyte Cells (male)	No. 157	Similar to signal observed for probe no. 161.

c. Southern blot hybridization using isolated segments of the megachromosome as probes

- DNA was isolated from mouse spleen tissue, mouse LMTK⁻ cells, 5 K20 Chinese hamster ovary cells, EJ30 human fibroblast cells and H1D3 cells. The isolated DNA and lambda phage DNA, was subjected to Southern blot hybridization using inserts isolated from megachromosome clone nos. 30, 113, 157 and 161 as probes. Plasmid pWE15 was used as a negative control probe. Each of the four megachromosome clone 10 inserts hybridized in a multi-copy manner (as demonstrated by the intensity of hybridization and the number of hybridizing bands) to all of the DNA samples, except the lambda phage DNA. Plasmid pWE15 hybridized to lambda DNA only.

d. Sequence analysis of megachromosome clone no. 161

- 15 Megachromosome clone no. 161 appeared to show the strongest hybridization in the *in situ* and Southern hybridization experiments and was chosen for analysis of the insert sequence. The sequence analysis was approached by first subcloning the insert of cosmid clone no. 161 to obtain five subclones as follows.

- 20 To obtain the end fragments of the insert of clone no. 161, the clone was digested with NotI and BamHI and ligated with NotI/BamHI-digested pBluescript KS (Stratagene, La Jolla, California). Two fragments of the insert of clone no. 161 were obtained: a 0.2-kb and a 0.7-kb insert

fragment. To subclone the internal fragment of the insert of clone no. 161, the same digest was ligated with BamHI-digested pUC19. Three fragments of the insert of clone no. 161 were obtained: a 0.6-kb, a 1.8-kb and a 4.8-kb insert fragment.

- 5 The ends of all the subcloned insert fragments were first sequenced manually. However, due to their extremely high GC content, autoradiographs were difficult to interpret and sequencing was repeated using an ABI sequencer and the dye-terminator cycle protocol. A comparison of the sequence data to sequences in the GENBANK database
- 10 revealed that the insert of clone no. 161 corresponds to an internal section of the mouse ribosomal RNA gene (rDNA) repeat unit between positions 7551-15670 as set forth in GENBANK accession no. X82564, which is provided as SEQ ID NO. 16 herein. The sequence data obtained for the insert of clone no. 161 is set forth in SEQ ID NOS. 18-24.
- 15 Specifically, the individual subclones corresponded to the following positions in GENBANK accession no. X82564 (i.e., SEQ ID NO. 16) and in SEQ ID NOS. 18-24:

20

Subclone	Start	End	Site	SEQ ID No.
	in X82564			
161k1	7579	7755	<u>Not</u> I, <u>Bam</u> HI	18
161m5	7756	8494	<u>Bam</u> HI	19
161m7	8495	10231	<u>Bam</u> HI	20 (shows only sequence corresponding to nt. 8495-8950), 21 (shows only sequence corresponding to nt. 9851- 10231)
161m12	10232	15000	<u>Bam</u> HI	22 (shows only sequence corresponding to nt. 10232-10600), 23 (shows only sequence corresponding to nt. 14267-15000),
25 161k2	15001	15676	<u>Not</u> I, <u>Bam</u> HI	24

The sequence set forth in SEQ ID NOS. 18-24 diverges in some positions from the sequence presented in positions 7551-15670 of GENBANK accession no. X82564. Such divergence may be attributable

- to random mutations between repeat units of rDNA. The results of the sequence analysis of clone no. 161, which reveal that it corresponds to rDNA, correlate with the appearance of the *in situ* hybridization signal it generated in human lymphocytes and mouse spleen cells. The
- 5 hybridization signal was clearly observed on acrocentric chromosomes in these cells, and such types of chromosomes are known to include rDNA adjacent to the pericentric satellite DNA on the short arm of the chromosome. Furthermore, rRNA genes are highly conserved in mammals as supported by the cross-species hybridization of clone no. 161 to
- 10 human chromosomal DNA.

- To isolate amplification-replication control regions such as those found in rDNA, it may be possible to subject DNA isolated from megachromosome-containing cells, such as H1D3 cells, to nucleic acid amplification using, e.g., the polymerase chain reaction (PCR) with the
- 15 following primers:
- amplification control element forward primer (1-30)
 5'-GAGGAATTCCCCATCCCTAATCCAGATTGGTG-3' (SEQ ID NO. 25)
 amplification control element reverse primer (2142-2111)
 5'-AAACTGCAGGCCGAGCCACCTCTCTTGTGTGTTG-3' (SEQ ID NO. 26)
- 20 origin of replication region forward primer (2116-2141)
 5'-AGGAATTCACAGAAGAGAGGTGGCTCGGCCTGC-3' (SEQ ID NO. 27)
 origin of replication region reverse primer (5546-5521)
 5'-AGCCTGCAGGAAGTCATACCTGGGAGGTGGCCC-3' (SEQ ID NO. 28)

C. Summary of the formation of the megachromosome

- 25 Figure 2 schematically sets forth events leading to the formation of a stable megachromosome beginning with the generation of a dicentric chromosome in a mouse LMTK⁻ cell line: (A) A single E-type amplification in the centromeric region of the mouse chromosome 7 following transfection of LMTK⁻ cells with λ CM8 and λ gtWESneo generates the neo-
- 30 centromere linked to the integrated foreign DNA, and forms a dicentric chromosome. Multiple E-type amplification forms the λ neo-chromosome,

- which was derived from chromosome 7 and stabilized in a mouse-hamster hybrid cell line; (B) Specific breakage between the centromeres of a dicentric chromosome 7 generates a chromosome fragment with the neo-centromere, and a chromosome 7 with traces of foreign DNA at the end;
- 5 (C) Inverted duplication of the fragment bearing the neo-centromere results in the formation of a stable neo-minichromosome; (D) Integration of exogenous DNA into the foreign DNA region of the formerly dicentric chromosome 7 initiates H-type amplification, and the formation of a heterochromatic arm. By capturing a euchromatic terminal segment, this
- 10 new chromosome arm is stabilized in the form of the "sausage" chromosome; (E) BrdU treatment and/or drug selection appears to induce further H-type amplification, which results in the formation of an unstable gigachromosome; (F) Repeated BrdU treatments and/or drug selection induce further H-type amplification including a centromere duplication,
- 15 which leads to the formation of another heterochromatic chromosome arm. It is split off from the chromosome 7 by chromosome breakage and acquires a terminal segment to form the stable megachromosome.

D. Expression of β -galactosidase and hygromycin transferase genes in cell lines carrying the megachromosome or derivatives thereof

- 20 The level of heterologous gene (i.e., β -galactosidase and hygromycin transferase genes) expression in cell lines containing the megachromosome or a derivative thereof was quantitatively measured. The relationship between the copy-number of the heterologous genes and the level of protein expressed therefrom was also determined.

25 1. Materials and methods

a. Cell lines

- Heterologous gene expression levels of H1D3 cells, carrying a 250-400 Mb megachromosome as described above, and mM2C1 cells, carrying a 50-60 Mb micro-megachromosome, were quantitatively evaluated.
- 30 mM2C1 cells were generated by repeated BrdU treatment and single cell cloning of the H1xHe41 cell line (mouse-hamster-human hybrid cell line

carrying the megachromosome and a single human chromosome with CD4 and neo^r genes; see Figure 4). The cell lines were grown under standard conditions in F12 medium under selective (120 μ g/ml hygromycin) or non-selective conditions.

5 b. Preparation of cell extract for β -galactosidase assays

Monolayers of mM2C1 or H1D3 cell cultures were washed three times with phosphate-buffered saline (PBS). Cells were scraped by rubber policemen and suspended and washed again in PBS. Washed cells were resuspended into 0.25 M Tris-HCl, pH 7.8, and disrupted by three cycles
10 of freezing in liquid nitrogen and thawing at 37°C. The extract was clarified by centrifugation at 12,000 rpm for 5 min. at 4°C.

c. β -galactosidase assay

The β -galactosidase assay mixture contained 1 mM MgCl₂, 45 mM β -mercaptoethanol, 0.8 mg/ml o-nitrophenyl- β -D-galactopyranoside and
15 66 mM sodium phosphate, pH 7.5. After incubating the reaction mixture with the cell extract at 37°C for increasing time, the reaction was terminated by the addition of three volumes of 1M Na₂CO₃, and the optical density was measured at 420 nm. Assay mixture incubated without cell extract was used as a control. The linear range of the reaction
20 was determined to be between 0.1-0.8 OD₄₂₀. One unit of β -galactosidase activity is defined as the amount of enzyme that will hydrolyse 3 nmoles of o-nitrophenyl- β -D-galactopyranoside in 1 minute at 37°C.

d. Preparation of cell extract for hygromycin phosphotransferase assay

25 Cells were washed as described above and resuspended into 20 mM Hepes buffer, pH 7.3, 100 mM potassium acetate, 5 mM Mg acetate and 2 mM dithiothreitol). Cells were disrupted at 0°C by six 10 sec bursts in an MSE ultrasonic disintegrator using a microtip probe. Cells were allowed to cool for 1 min after each ultrasonic burst. The extracts
30 were clarified by centrifuging for 1 min at 2000 rpm in a microcentrifuge.

e. **Hygromycin phosphotransferase assay**

Enzyme activity was measured by means of the phosphocellulose paper binding assay as described by Haas and Dowding [(1975). Meth. Enzymol. 43:611-628]. The cell extract was supplemented with 0.1 M ammonium chloride and 1 mM adenosine- γ -³²P-triphosphate (specific activity: 300 Ci/mmol). The reaction was initiated by the addition of 0.1 mg/ml hygromycin and incubated for increasing time at 37°C. The reaction was terminated by heating the samples for 5 min at 75°C in a water bath, and after removing the precipitated proteins by centrifugation for 5 min in a microcentrifuge, an aliquot of the supernatant was spotted on a piece of Whatman P-81 phosphocellulose paper (2 cm²). After 30 sec at room temperature the papers are placed into 500 ml of hot (75°C) distilled water for 3 min. While the radioactive ATP remains in solution under these conditions, hygromycin phosphate binds strongly and quantitatively to phosphocellulose. The papers are rinsed 3 times in 500 ml of distilled water and the bound radioactivity was measured in toluene scintillation cocktail in a Beckman liquid scintillation counter. Reaction mixture incubated without added hygromycin served as a control.

f. **Determination of the copy-number of the heterologous genes**

DNA was prepared from the H1D3 and mM2C1 cells using standard purification protocols involving SDS lysis of the cells followed by Proteinase K treatment and phenol/chloroform extractions. The isolated DNA was digested with an appropriate restriction endonuclease, fractionated on agarose gels, blotted to nylon filters and hybridized with a radioactive probe derived either from the β -galactosidase or the hygromycin phosphotransferase genes. The level of hybridization was quantified in a Molecular Dynamics PhosphorImage Analyzer. To control the total amount of DNA loaded from the different cells lines, the filters were reprobed with a single copy gene, and the hybridization of β -

galactosidase and hygromycin phosphotransferase genes was normalized to the single copy gene hybridization.

g. Determination of protein concentration

The total protein content of the cell extracts was measured by the

- 5 Bradford colorimetric assay using bovine serum albumin as standard.

2. Characterization of the β -galactosidase and hygromycin phosphotransferase activity expressed in H1D3 and mM2C1 cells

In order to establish quantative conditions, the most important

- 10 kinetic parameters of β -galactosidase and hygromycin phosphotransferase activity have been studied. The β -galactosidase activity measured with a colorimetric assay was linear between the 0.1-0.8 OD₄₂₀ range both for the mM2C1 and H1D3 cell lines. The β -galactosidase activity was also proportional in both cell lines with the amount of protein added to the
- 15 reaction mixture within 5-100 μ g total protein concentration range. The hygromycin phosphotransferase activity of mM2C1 and H1D3 cell lines was also proportional with the reaction time or the total amount of added cell extract under the conditions described for the β -galactosidase.

- 20 **a. Comparison of β -galactosidase activity of mM2C1 and H1D3 cell lines**

- Cell extracts prepared from logarithmically growing mM2C1 and H1D3 cell lines were tested for β -galactosidase activity, and the specific activities were compared in 10 independent experiments. The β -galactosidase activity of H1D3 cell extracts was 440 ± 25 U/mg total
- 25 protein. Under identical conditions the β -galactosidase activity of the mM2C1 cell extracts was 4.8 times lower: 92 ± 13 U/mg total protein.

- β -galactosidase activities of highly subconfluent, subconfluent and nearly confluent cultures of H1D3 and mM2C1 cell lines were also compared. In these experiments different numbers of logarithmic H1D3
- 30 and mM2C1 cells were seeded in constant volume of culture medium and grown for 3 days under standard conditions. No significant difference was found in the β -galactosidase specific activities of cell cultures grown

at different cell densities, and the ratio of H1D3/mM2C1 β -galactosidase specific activities was also similar for all three cell densities. In confluent, stationary cell cultures of H1D3 or mM2C1 cells, however, the expression of β -galactosidase significantly decreased due likely to cessation of cell

5 division as a result of contact inhibition.

b. Comparison of hygromycin phosphotransferase activity of H1D3 and mM2C1 cell lines

The bacterial hygromycin phosphotransferase is present in a membrane-bound form in H1D3 or mM2C1 cell lines. This follows from
10 the observation that the hygromycin phosphotransferase activity can be completely removed by high speed centrifugation of these cell extracts, and the enzyme activity can be recovered by resuspending the high speed pellet.

The ratio of the enzyme's specific activity in H1D3 and mM2C1 cell
15 lines was similar to that of β -galactosidase activity, i.e., H1D3 cells have 4.1 times higher specific activity compared with mM2C1 cells.

c. Hygromycin phosphotransferase activity in H1D3 and mM2C1 cells grown under non-selective conditions

The level of expression of the hygromycin phosphotransferase gene
20 was measured on the basis of quantitation of the specific enzyme activities in H1D3 and mM2C1 cell lines grown under non-selective conditions for 30 generations. The absence of hygromycin in the medium did not influence the expression of the hygromycin phosphotransferase gene.

25 3. Quantitation of the number of β -galactosidase and hygromycin phosphotransferase gene copies in H1D3 and mM2C1 cell lines

As described above, the β -galactosidase and hygromycin
30 phosphotransferase genes are located only within the megachromosome, or micro-megachromosome in H1D3 and mM2C1 cells. Quantitative analysis of genomic Southern blots of DNA isolated from H1D3 and mM2C1 cell lines with the PhosphorImage Analyzer revealed that the copy number

of β -galactosidase genes integrated into the megachromosome is approximately 10 times higher in H1D3 cells than in mM2C1 cells. The copy-number of hygromycin phosphotransferase genes is approximately 7 times higher in H1D3 cells than in mM2C1 cells.

5 4. **Summary and conclusions of results of quantitation of heterologous gene expression in cells containing megachromosomes or derivatives thereof**

Quantitative determination of β -galactosidase activity of higher eukaryotic cells (e.g., H1D3 cells) carrying the bacterial β -galactosidase gene in heterochromatic megachromosomes confirmed the observed high-level expression of the integrated bacterial gene detected by cytological staining methods. It has generally been established in reports of studies of the expression of foreign genes in transgenic animals that, although transgene expression shows correct tissue and developmental specificity, 10 the level of expression is typically low and shows extensive position-dependent variability (i.e., the level of transgene expression depends on the site of chromosomal integration). It has been assumed that the low-level transgene expression may be due to the absence of special DNA sequences which can insulate the transgene from the inhibitory effect of 15 the surrounding chromatin and promote the formation of active chromatin structure required for efficient gene expression. Several cis-activating DNA sequence elements have been identified that abolish this position-dependent variability, and can ensure high-level expression of the transgene locus activating region (LAR) sequences in higher eukaryotes and specific 20 chromatin structure (scs) elements in lower eukaryotes (see, et al. Eisenberg and Elgin (1991) Trends in Genet. 7:335-340). If these cis-acting DNA sequences are absent, the level of transgene expression is low and copy-number independent. 25

Although the bacterial β -galactosidase reporter gene contained in 30 the heterochromatic megachromosomes of H1D3 and mM2C1 cells is driven by a potent eukaryotic promoter-enhancer element, no specific cis-acting DNA sequence element was designed and incorporated into the

bacterial DNA construct which could function as a boundary element. Thus, the high-level β -galactosidase expression measured in these cells is of significance, particularly because the β -galactosidase gene in the megachromosome is located in a long, compact heterochromatic environment, which is known to be able to block gene expression. The megachromosome appears to contain DNA sequence element(s) in association with the bacterial DNA sequences that function to override the inhibitory effect of heterochromatin on gene expression.

The specificity of the heterologous gene expression in the megachromosome is further supported by the observation that the level of β -galactosidase expression is copy-number dependent. In the H1D3 cell line, which carries a full-size megachromosome, the specific activity of β -galactosidase is about 5-fold higher than in mM2C1 cells, which carry only a smaller, truncated version of the megachromosome. A comparison of the number of β -galactosidase gene copies in H1D3 and mM2C1 cell lines by quantitative hybridization techniques confirmed that the expression of β -galactosidase is copy-number dependent. The number of integrated β -galactosidase gene copies is approximately 10-fold higher in the H1D3 cells than in mM2C1 cells. Thus, the cell line containing the greater number of copies of the β -galactosidase gene also yields higher levels of β -galactosidase activity, which supports the copy-number dependency of expression. The copy number dependency of the β -galactosidase and hygromycin phosphotransferase enzyme levels in cell lines carrying different derivatives of the megachromosome indicates that neither the chromatin organization surrounding the site of integration of the bacterial genes, nor the heterochromatic environment of the megachromosome suppresses the expression of the genes.

The relative amount of β -galactosidase protein expressed in H1D3 cells can be estimated based on the V_{\max} of this enzyme [500 for homogeneous, crystallized bacterial β -galactosidase (Naider et al. (1972) Biochemistry 11:3202-3210)] and the specific activity of H1D3 cell

- protein. A V_{\max} of 500 means that the homogeneous β -galactosidase protein hydrolyzes 500 μ moles of substrate per minute per mg of enzyme protein at 37°C. One mg of total H1D3 cell protein extract can hydrolyze 1.4 μ moles of substrate per minute at 37°C, which means that 0.28% of
- 5 the protein present in the H1D3 cell extract is β -galactosidase.

- The hygromycin phosphotransferase is present in a membrane-bound form in H1D3 and mM2C1 cells. The tendency of the enzyme to integrate into membranes in higher eukaryotic cells may be related to its periplasmic localization in prokaryotic cells. The bacterial hygromycin
- 10 phosphotransferase has not been purified to homogeneity; thus, its V_{\max} has not been determined. Therefore, no estimate can be made on the total amount of hygromycin phosphotransferase protein expressed in these cell lines. The 4-fold higher specific activity of hygromycin phosphotransferase in H1D3 cells as compared to mM2C1 cells, however,
- 15 indicates that its expression is also copy number dependent.

- The constant and high level expression of the β -galactosidase gene in H1D3 and mM2C1 cells, particularly in the absence of any selective pressure for the expression of this gene, clearly indicates the stability of the expression of genes carried in the heterochromatic megachromo-
- 20 somes. This conclusion is further supported by the observation that the level of hygromycin phosphotransferase expression did not change when H1D3 and mM2C1 cells were grown under non-selective conditions. The consistent high-level, stable, and copy-number dependent expression of bacterial marker genes clearly indicates that the megachromosome is an
- 25 ideal vector system for expression of foreign genes.

EXAMPLE 7

Summary of some of the cell lines with SATACS and minichromosomes that have been constructed

30 1. EC3/7-Derived cell lines

The LMTK'-derived cell line, which is a mouse fibroblast cell line, was transfected with λ CM8 and λ gtWESneo DNA [see, EXAMPLE 2] to

produce transformed cell lines. Among these, was was EC3/7, deposited at the European Collection of Animal cell Culture (ECACC) under Accession No. 90051001 [see, U.S. Patent No. 5,288,625; see, also Hadlaczký et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110 and U.S.

- 5 application Serial No. 08/375,271]. This cell line contains the dicentric chromosome with the neo-centromere. Recloning and selection produced cell lines such as EC3/7C5, which are cell lines with the stable neo-minichromosome and the formerly dicentric chromosome [see, Fig. 2C].

2. KE1-2/4 Cells

- 10 Fusion of EC3/7 with CHO-K20 cells and selection with G418/HAT produced hybrid cell lines, among these was KE1-2/4, which has been deposited with the ECACC under Accession No. 96040924. KE1-2/4 is a stable cell line that contains the λ neo-chromosome [see, Fig. 2D; see, also U.S. Patent No. 5,288,625], produced by E-type amplifications.
- 15 KE1-2/4 has been transfected with vectors containing λ DNA, selectable markers, such as the puromycin-resistance gene, and genes of interest, such as p53 and the anti-HIV ribozyme gene. These vectors target the gene of interest into the λ neo-chromosome by virtue of homologous recombination with the heterologous DNA in the chromosome.

20 3. C5pMCT53 Cells

The EC3/7C5 cell line has been co-transfected with pH132, pCH110 and λ DNA [see, EXAMPLE 2] as well as other constructs. Various clones and subclones have been selected. For example transformation with a construct that includes p53 encoding DNA,

- 25 produced cells designated C5pMCT53.

4. TF1004G24 Cells

As discussed above, cotransfection of EC3/7C5 cells with plasmids [pH132, pCH110 available from Pharmacia, see, also Hall et al. (1983) J. Mol. Appl. Gen. 2:101-109] and with λ DNA [λ cl 875 Sam 7 (New England Biolabs)] produced transformed cells. Among these is TF1004G24, which contains the DNA encoding the anti-HIV ribozyme in the neo-mini-

chromosome. Recloning of TF1004G24 produced numerous cell lines. Among these is the NHHL24 cell line. This cell line also has the anti-HIV ribozyme in the neo-minichromosome and expresses high levels of β -gal. It has been fused with CHO-K20 cells to produce various hybrids.

5. TF1004G19-Derived cells

- Recloning and selection of the TF1004G transformants produced the cell line TF1004G19, discussed above in EXAMPLE 4, which contains the unstable sausage chromosome and the neo-minichromosome. Single cell cloning produced the TF1004G-19C5 [see Figure 4] cell line, which has a stable sausage chromosome and the neo-minichromosome. TF1004G-19C5 has been fused with CHO cells and the hybrids grown under selective conditions to produce the 19C5xHa4 and 19C5xHa3 cell lines [see, EXAMPLE 4] and others. Recloning of the 19C5xHa3 cell line yielded a cell line containing a gigachromosome, i.e., cell line 19C5xHa47, see Figure 2E. BrdU treatment of 19C5xHa4 cells and growth under selective conditions [neomycin (G) and/or hygromycin (H)] has produced hybrid cell lines such as the G3D5 and G4D6 cell lines and others. G3D5 has the neo-minichromosome and the megachromosome. G4D6 has only the neo-minichromosome.
- Recloning of 19C5xHa4 cells in H medium produced numerous clones. Among these is H1D3 [see Figure 4], which has the stable megachromosome. Repeated BrdU treatment and recloning of H1D3 cells has produced the HB31 cell line, which has been used for transformations with the pTEMPUD, pTEMPU, pTEMPU3, and pCEPUR-132 vectors [see, Examples 12 and 14, below].

- H1D3 has been fused with a CD4⁺ Hela cell line that carries DNA encoding CD4 and neomycin resistance on a plasmid [see, e.g., U.S. Patent Nos. 5,413,914, 5,409,810, 5,266,600, 5,223,263, 5,215,914 and 5,144,019, which describe these Hela cells]. Selection with GH has produced hybrids, including H1xHE41 [see Figure 4], which carries the megachromosome and also a single human chromosome that includes the

- CD4neo construct. Repeated BrdU treatment and single cell cloning has produced cell lines with the megachromosome [cell line 1B3, see Figure 4]. About 25% of the 1B3 cells have a truncated megachromosome [~90-120 Mb]. Another of these subclones, designated 2C5, was
- 5 cultured on hygromycin-containing medium and megachromosome-free cell lines were obtained and grown in G418-containing medium. Recloning of these cells yielded cell lines such as 1B4 and others that have a dwarf megachromosome [~150-200 Mb], and cell lines, such as 11C3 and mM2C1, which have a micro-megachromosome [~50-90 Mb]. The
- 10 micro-megachromosome of cell line mM2C1 has no telomeres; however, if desired, synthetic telomeres, such as those described and generated herein, may be added to the mM2C1 cell micro-megachromosomes. Cell lines containing smaller truncated megachromosomes, such as the mM2C1 cell line containing the micro-megachromosome, can be used to
- 15 generate even smaller megachromosomes, e.g., ~10-30 Mb in size. This may be accomplished, for example, by breakage and fragmentation of the micro-megachromosome in these cells through exposing the cells to X-ray irradiation, BrdU or telomere-directed in vivo chromosome fragmentation.

EXAMPLE 8

20 Replication of the megachromosome

The homogeneous architecture of the megachromosomes provides a unique opportunity to perform a detailed analysis of the replication of the constitutive heterochromatin.

A. Materials and methods

25 1. Culture of cell lines

- H1D3 mouse-hamster hybrid cells carrying the megachromosome [see, EXAMPLE 4] were cultured in F-12 medium containing 10% fetal calf serum [FCS] and 400 μ g/ml Hygromycin B [Calbiochem]. G3D5 hybrid cells [see, Example 4] were maintained in F-12 medium containing
- 30 10% FCS, 400 μ g/ml Hygromycin B [Calbiochem], and 400 μ g/ml G418

[SIGMA]. Mouse A9 fibroblast cells were cultured in F-12 medium supplemented with 10% FCS.

2. BrdU labelling

In typical experiments, 20-24 parallel semi-confluent cell cultures were set up in 10 cm Petri dishes. Bromodeoxyuridine (BrdU) (Fluka) was dissolved in distilled water alkalized with a drop of NaOH, to make a 10^{-2} M stock solution. Aliquots of 10-50 μ l of this BrdU stock solution were added to each 10 ml culture, to give a final BrdU concentration of 10-50 μ M. The cells were cultured in the presence of BrdU for 30 min, and then washed with warm complete medium, and incubated without BrdU until required. At this point, 5 μ g/ml colchicine was added to a sample culture every 1 or 2 h. After 1-2 h colchicine treatment, mitotic cells were collected by "shake-off" and regular chromosome preparations were made for immunolabelling.

3. Immunolabelling of chromosomes and *in situ* hybridization

Immunolabelling with fluorescein-conjugated anti-BrdU monoclonal antibody (Boehringer) was done according to the manufacturer's recommendations, except that for mouse A9 chromosomes, 2 M hydrochloric acid was used at 37° C for 25 min, while for chromosomes of hybrid cells, 1 M hydrochloric acid was used at 37° C for 30 min. *In situ* hybridization with biotin-labelled probes, and indirect immunofluorescence and *in situ* hybridization on the same preparation, were performed as described previously [Hadlaczký *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:8106-8110, see, also U.S. Patent No. 5,288,625].

4. Microscopy

All observations and microphotography were made by using a Vanox AHBS (Olympus) microscope. Fujicolor 400 Super G or Fujicolor 1600 Super HG high-speed colour negatives were used for photographs.

B. Results

The replication of the megachromosome was analyzed by BrdU pulse labelling followed by immunolabelling. The basic parameters for DNA labelling *in vivo* were first established. Using a 30-min pulse of 50 μ M BrdU in parallel cultures, samples were taken and fixed at 5 min intervals from the beginning of the pulse, and every 15 min up to 1 h after the removal of BrdU. Incorporated BrdU was detected by immunolabelling with fluorescein-conjugated anti-BrdU monoclonal antibody. At the first time point (5 min) 38% of the nuclei were labelled, and a gradual increase in the number of labelled nuclei was observed during incubation in the presence of BrdU, culminating in 46% in the 30-min sample, at the time of the removal of BrdU. At further time points (60, 75, and 90 min) no significant changes were observed, and the fraction of labelled nuclei remained constant [44.5-46%].

These results indicate that (i) the incorporation of the BrdU is a rapid process, (ii) the 30 min pulse-time is sufficient for reliable labelling of S-phase nuclei, and (iii) the BrdU can be effectively removed from the cultures by washing.

The length of the cell cycle of the H1D3 and G3D5 cells was estimated by measuring the time between the appearance of the earliest BrdU signals on the extreme late replicating chromosome segments and the appearance of the same pattern only on one of the chromatids of the chromosomes after one completed cell cycle. The length of G2 period was determined by the time of the first detectable BrdU signal on prophase chromosomes and by the labelled mitoses method [Qastler *et al.* (1959) *Exp. Cell Res.* 17:420-438]. The length of the S-phase was determined in three ways: (i) on the basis of the length of cell cycle and the fraction of nuclei labelled during the 30-120 min pulse; (ii) by measuring the time between the very end of the replication of the extreme late replicating chromosomes and the detection of the first signal on the chromosomes at the beginning of S phase; (iii) by the labelled

mitoses method. In repeated experiments, the duration of the cell cycle was found to be 22-26 h, the S phase 10-14 h, and the G2 phase 3.5-4.5 h.

- Analyses of the replication of the megachromosome were made in parallel cultures by collecting mitotic cells at two hour intervals following two hours of colchicine treatment. In a repeat experiment, the same analysis was performed using one hour sample intervals and one hour colchicine treatment. Although the two procedures gave comparable results, the two hour sample intervals were viewed as more appropriate since approximately 30% of the cells were found to have a considerably shorter or longer cell cycle than the average. The characteristic replication patterns of the individual chromosomes, especially some of the late replicating hamster chromosomes, served as useful internal markers for the different stages of S-phase. To minimize the error caused by the different lengths of cell cycles in the different experiments, samples were taken and analyzed throughout the whole cell cycle until the appearance of the first signals on one chromatid at the beginning of the second S-phase.

The sequence of replication in the megachromosome is as follows.

- At the very beginning of the S-phase, the replication of the megachromosome starts at the ends of the chromosomes. The first initiation of replication in an interstitial position can usually be detected at the centromeric region. Soon after, but still in the first quarter of the S-phase, when the terminal region of the short arm has almost completed its replication, discrete initiation signals appear along the chromosome arms. In the second quarter of the S-phase, as replication proceeds, the BrdU-labelled zones gradually widen, and the checkered pattern of the megachromosome becomes clear [see, e.g., Fig. 2F]. At the same time, pericentric regions of mouse chromosomes also show intense incorporation of BrdU. The replication of the megachromosome peaks at the end of the second quarter and in the third quarter of the S-phase. At

the end of the third quarter, and at the very beginning of the last quarter of the S-phase, the megachromosome and the pericentric heterochromatin of the mouse chromosomes complete their replication. By the end of S-phase, only the very late replicating segments of mouse and hamster

5 chromosomes are still incorporating BrdU.

The replication of the whole genome occurs in distinct phases. The signal of incorporated BrdU increased continuously until the end of the first half of the S-phase, but at the beginning of the third quarter of the S-phase chromosome segments other than the heterochromatic regions

10 hardly incorporated BrdU. In the last quarter of the S-phase, the BrdU signals increased again when the extreme late replicating segments showed very intense incorporation.

Similar analyses of the replication in mouse A9 cells were performed as controls. To increase the resolution of the immunolabelling

15 pattern, pericentric regions of A9 chromosomes were decondensed by treatment with Hoechst 33258. Because of the intense replication of the surrounding euchromatic sequences, precise localization of the initial BrdU signal in the heterochromatin was normally difficult, even on

20 undercondensed mouse chromosomes. On those chromosomes where the initiation signal(s) were localized unambiguously, the replication of the pericentric heterochromatin of A9 chromosomes was similar to that of the megachromosome. Chromosomes of A9 cells also exhibited replication patterns and sequences similar to those of the mouse chromosomes in the hybrid cells. These results indicate that the replicators of the

25 megachromosome and mouse chromosomes retained their original timing and specificity in the hybrid cells.

By comparing the pattern of the initiation sites obtained after BrdU incorporation with the location of the integration sites of the "foreign" DNA in a detailed analysis of the first quarter of the S-phase, an attempt

30 was made to identify origins of replication (initiation sites) in relation to the amplicon structure of the megachromosome. The double band of

integrated DNA on the long arm of the megachromosome served as a cytological marker. The results showed a colocalization of the BrdU and *in situ* hybridization signals found at the cytological level, indicating that the "foreign" DNA sequences are in close proximity to the origins of

- 5 replication, presumably integrated into the non-satellite sequences between the replicator and the satellite sequences [see, Figure 3]. As described in Example 6.B.4, the rDNA sequences detected in the megachromosome are also localized at the amplicon borders at the site of integration of the "foreign" DNA sequences, suggesting that the origins of
- 10 replication responsible for initiation of replication of the megachromosome involve rDNA sequences. In the pericentric region of several other chromosomes, dot-like BrdU signals can also be observed that are comparable to the initiation signals on the megachromosome. These signals may represent similar initiation sites in the heterochromatic
- 15 regions of normal chromosomes.

- At a frequency of 10^{-4} , "uncontrolled" amplification of the integrated DNA sequences was observed in the megachromosome. Consistent with the assumption (above) that "foreign" sequences are in proximity of the replicators, this spatially restricted amplification is likely
- 20 to be a consequence of uncontrolled repeated firings of the replication origin(s) without completing the replication of the whole segment.

C. Discussion

- It has generally been thought that the constitutive heterochromatin of the pericentric regions of chromosomes is late replicating [see, e.g.,
- 25 Miller (1976) Chromosoma 55:165-170]. On the contrary, these experiments evidence that the replication of the heterochromatic blocks starts at a discrete initiation site in the first half of the S-phase and continues through approximately three-quarters of S-phase. This difference can be explained in the following ways: (i) in normal
- 30 chromosomes, actively replicating euchromatic sequences that surround the satellite DNA obscure the initiation signals, and thus the precise

localization of initiation sites is obscured; (ii) replication of the heterochromatin can only be detected unambiguously in a period during the second half of the S-phase, when the bulk of the heterochromatin replicates and most other chromosomal regions have already completed their replication, or have not yet started it. Thus, low resolution cytological techniques, such as analysis of incorporation of radioactively labelled precursors by autoradiography, only detect prominent replication signals in the heterochromatin in the second half of S-phase, when adjacent euchromatic segments are no longer replicating.

- 10 In the megachromosome, the primary initiation sites of replication colocalize with the sites where the "foreign" DNA sequences and rDNA sequences are integrated at the amplicon borders. Similar initiation signals were observed at the same time in the pericentric heterochromatin of some of the mouse chromosomes that do not have "foreign" DNA,
- 15 indicating that the replication initiation sites at the borders of amplicons may reside in the non-satellite flanking sequences of the satellite DNA blocks. The presence of a primary initiation site at each satellite DNA doublet implies that this large chromosome segment is a single huge unit of replication [megareplicon] delimited by the primary initiation site and
- 20 the termination point at each end of the unit. Several lines of evidence indicate that, within this higher-order replication unit, "secondary" origins and replicons contribute to the complete replication of the megareplicon:

1. The total replication time of the heterochromatic regions of the megachromosome was ~9-11 h. At the rate of movement of replication forks, 0.5-5 kb per minute, that is typical of eukaryotic chromosomes [Kornberg et al. (1992) *DNA Replication*. 2nd. ed., New York: W.H. Freeman and Co, p. 474], replication of a ~15 Mb replicon would require 50-500 h. Alternatively, if only a single replication origin was used, the average replication speed would have to be 25 kb per
- 30 minute to complete replication within 10 h. By comparing the intensity of the BrdU signals on the euchromatic and the heterochromatic

chromosome segments, no evidence for a 5- to 50-fold difference in their replication speed was found.

2. Using short BrdU pulse labelling, a single origin of replication would produce a replication band that moves along the replicon, reflecting the movement of the replication fork. In contrast, a widening of the replication zone that finally gave rise to the checkered pattern of the megachromosome was observed, and within the replication period, the most intensive BrdU incorporation occurred in the second half of the S-phase. This suggests that once the megareplicator has been activated, it permits the activation and firing of "secondary" origins, and that the replication of the bulk of the satellite DNA takes place from these "secondary" origins during the second half of the S-phase. This is supported by the observation that in certain stages of the replication of the megachromosome, the whole amplicon can apparently be labelled by a short BrdU pulse.

- Megareplicators and secondary replication origins seem to be under strict temporal and spatial control. The first initiation within the megachromosomes usually occurred at the centromere, and shortly afterward all the megareplicators become active. The last segment of the megachromosome to complete replication was usually the second segment of the long arm. Results of control experiments with mouse A9 chromosomes indicate that replication of the heterochromatin of mouse chromosomes corresponds to the replication of the megachromosome amplicons. Therefore, the pre-existing temporal control of replication in the heterochromatic blocks is preserved in the megachromosome.
- Positive [Hassan *et al.* (1994) *J. Cell. Sci.* **107**:425-434] and negative [Haase *et al.* (1994) *Mol. Cell. Biol.* **14**:2516-2524] correlations between transcriptional activity and initiation of replication have been proposed. In the megachromosome, transcription of the integrated genes seems to have no effect on the original timing of the replication origins. The concerted, precise timing of the megareplicator initiations in the different

amplicons suggests the presence of specific, cis-acting sequences, origins of replication.

Considering that pericentric heterochromatin of mouse chromosomes contains thousands of short, simple repeats spanning 7-
5 15 Mb, and the centromere itself may also contain hundreds of kilobases, the existence of a higher-order unit of replication seems probable. The observed uncontrolled intrachromosomal amplification restricted to a replication initiation region of the megachromosome is highly suggestive of a rolling-circle type amplification, and provides additional evidence for
10 the presence of a replication origin in this region.

The finding that a specific replication initiation site occurs at the boundaries of amplicons suggests that replication might play a role in the amplification process. These results suggest that each amplicon of the megachromosome can be regarded as a huge megareplicon defined by a
15 primary initiation site [megareplicator] containing "secondary" origins of replication. Fusion of replication bubbles from different origins of bi-directional replication [DePamphilis (1993) Ann. Rev. Biochem. 62:29-63] within the megareplicon could form a giant replication bubble, which would correspond to the whole megareplicon. In the light of this, the
20 formation of megabase-size amplicons can be accommodated by a replication-directed amplification mechanism. In H and E-type amplifications, intrachromosomal multiplication of the amplicons was observed [see, above EXAMPLES], which is consistent with the unequal sister chromatid exchange model. Induced or spontaneous unscheduled
25 replication of a megareplicon in the constitutive heterochromatin may also form new amplicon(s) leading to the expansion of the amplification or to the heterochromatic polymorphism of "normal" chromosomes. The "restoration" of the missing segment on the long arm of the megachromosome may well be the result of the re-replication of one
30 amplicon limited to one strand.

Taken together, without being bound by any theory, a replication-directed mechanism is a plausible explanation for the initiation of large-scale amplifications in the centromeric regions of mouse chromosomes, as well as for the *de novo* chromosome formations. If specific [amplificator, i.e., sequences controlling amplification] sequences play a role in promoting the amplification process, sequences at the primary replication initiation site [megareplicator] of the megareplicon are possible candidates.

- The presence of rRNA gene sequence at the amplicon borders near the foreign DNA in the megachromosome suggests that this sequence contributes to the primary replication initiation site and participates in large-scale amplification of the pericentric heterochromatin in *de novo* formation of SATACs. Ribosomal RNA genes have an intrinsic amplification mechanism that provides for multiple copies of tandem genes. Thus, for purposes herein, in the construction of SATACs in cells, rDNA will serve as a region for targeted integration, and as components of SATACs constructed in vitro.

EXAMPLE 9

Generation of chromosomes with amplified regions derived from mouse chromosome 1

- To show that the events described in EXAMPLES 2-7 are not unique to mouse chromosome 7 and to show that the EC7/3 cell line is not required for formation of the artificial chromosomes, the experiments have been repeated using different initial cell lines and DNA fragments. Any cell or cell line should be amenable to use or can readily be determined that it is not.

A. Materials

- The LP11 cell line was produced by the "scrape-loading" transfection method [Fechheimer et al. (1987) Proc. Natl. Acad. Sci. U.S.A. 84:8463-8467] using 25 μ g plasmid DNA for 5×10^6 recipient

cells. LP11 cells were maintained in F-12 medium containing 3-15 $\mu\text{g/ml}$ Puromycin [SIGMA].

B. Amplification in LP11 cells

- The large-scale amplification described in the above Examples is
- 5 not restricted to the transformed EC3/7 cell line or to the chromosome 7 of mouse. In an independent transformation experiment, LMTK⁻ cells were transfected using the calcium phosphate precipitation procedure with a selectable puromycin-resistance gene-containing construct designated pPuroTel [see Example 1.E.2. for a description of this plasmid], to
- 10 establish cell line LP11. Cell line LP11 carries chromosome(s) with amplified chromosome segments of different lengths [$\sim 150\text{-}600\text{ Mb}$]. Cytological analysis of the LP11 cells indicated that the amplification occurred in the pericentric region of the long arm of a submetacentric chromosome formed by Robertsonian translocation. This chromosome
- 15 arm was identified by G-banding as chromosome 1. C-banding and *in situ* hybridization with mouse major satellite DNA probe showed that an E-type amplification had occurred: the newly formed region was composed of an array of euchromatic chromosome segments containing different amounts of heterochromatin. The size and C-band pattern of the
- 20 amplified segments were heterogeneous. In several cells, the number of these amplified units exceeded 50; single-cell subclones of LP11 cell lines, however, carry stable marker chromosomes with 10-15 segments and constant C-band patterns.

- Sublines of the thymidine kinase-deficient LP11 cells (e.g., LP11-
- 25 15P 1C5/7 cell line) established by single-cell cloning of LP11 cells were transfected with a thymidine kinase gene construct. Stable TK⁺ transfectants were established.

EXAMPLE 10

Isolation of SATACS and other chromosomes with atypical base content and/or size

I. Isolation of artificial chromosomes from endogenous chromosomes

5

Artificial chromosomes, such as SATACs, may be sorted from endogenous chromosomes using any suitable procedures, and typically involve isolating metaphase chromosomes, distinguishing the artificial chromosomes from the endogenous chromosomes, and separating the artificial chromosomes from endogenous chromosomes. Such procedures will generally include the following basic steps: (1) culture of a sufficient number of cells (typically about 2×10^7 mitotic cells) to yield, preferably on the order of 1×10^6 artificial chromosomes, (2) arrest of the cell cycle of the cells in a stage of mitosis, preferably metaphase, using a mitotic arrest agent such as colchicine, (3) treatment of the cells, particularly by swelling of the cells in hypotonic buffer, to increase susceptibility of the cells to disruption, (4) by application of physical force to disrupt the cells in the presence of isolation buffers for stabilization of the released chromosomes, (5) dispersal of chromosomes in the presence of isolation buffers for stabilization of free chromosomes, (6) separation of artificial from endogenous chromosomes and (7) storage (and shipping if desired) of the isolated artificial chromosomes in appropriate buffers.

10

15

20

Modifications and variations of the general procedure for isolation of artificial chromosomes, for example to accommodate different cell types with differing growth characteristics and requirements and to optimize the duration of mitotic block with arresting agents to obtain the desired balance of chromosome yield and level of debris, may be empirically determined.

25

Steps 1-5 relate to isolation of metaphase chromosomes. The separation of artificial from endogenous chromosomes (step 6) may be accomplished in a variety of ways. For example, the chromosomes may be stained with DNA-specific dyes such as Hoeschst 33258 and

30

- chromomycin A₃ and sorted into artificial and endogenous chromosomes on the basis of dye content by employing fluorescence-activated cell sorting (FACS). To facilitate larger scale isolation of the artificial chromosomes, different separation techniques may be employed such as
- 5 swinging bucket centrifugation (to effect separation based on chromosome size and density) [see, e.g., Mendelsohn et al. (1968) J. Mol. Biol. 32:101-108], zonal rotor centrifugation (to effect separation on the basis of chromosome size and density) [see, e.g., Burki et al. (1973) Prep. Biochem. 3:157-182; Stubblefield et al. (1978) Biochem. Biophys. Res. Commun. 83:1404-1414, velocity sedimentation (to effect
 - 10 separation on the basis of chromosome size and shape) [see e.g., Collard et al. (1984) Cytometry 5:9-19]. Immuno-affinity purification may also be employed in larger scale artificial chromosome isolation procedures. In this process, large populations of artificial chromosome-containing cells
 - 15 (asynchronous or mitotically enriched) are harvested en masse and the mitotic chromosomes (which can be released from the cells using standard procedures such as by incubation of the cells in hypotonic buffer and/or detergent treatment of the cells in conjunction with physical
 - 20 disruption of the treated cells) are enriched by binding to antibodies that are bound to solid state matrices (e.g. column resins or magnetic beads). Antibodies suitable for use in this procedure bind to condensed centromeric proteins or condensed and DNA-bound histone proteins. For example, autoantibody LU851 (see Hadlaczký et al. (1989) Chromosoma 97:282-288), which recognizes mammalian centromeres may be used for
 - 25 large-scale isolation of chromosomes prior to subsequent separation of artificial from endogenous chromosomes using methods such as FACS. The bound chromosomes would be washed and eventually eluted for sorting. Immunoaffinity purification may also be used directly to separate artificial chromosomes from endogenous chromosomes. For example,
 - 30 SATACs may be generated in or transferred to (e.g., by microinjection or microcell fusion as described herein) a cell line that has chromosomes that

contain relatively small amounts of heterochromatin, such as hamster cells (e.g., V79 cells or CHO-K1 cells). The SATACs, which are predominantly heterochromatin, are then separated from the endogenous chromosomes by utilizing anti-heterochromatin binding protein (*Drosophila*

- 5 HP-1) antibody conjugated to a solid matrix. Such matrix preferentially binds SATACs relative to hamster chromosomes. Unbound hamster chromosomes are washed away from the matrix and the SATACs are eluted by standard techniques.

A. Cell lines and cell culturing procedures

- 10 In one isolation procedure, 1B3 mouse-hamster-human hybrid cells [see, Figure 4] carrying the megachromosome or the truncated megachromosome were grown in F-12 medium supplemented with 10% fetal calf serum, 150 $\mu\text{g/ml}$ hygromycin B and 400 $\mu\text{g/ml}$ G418. GHB42 [a cell line recloned from G3D5 cells] mouse-hamster hybrid cells carrying
- 15 the megachromosome and the minichromosome were also cultured in F-12 medium containing 10% fetal calf serum, 150 $\mu\text{g/ml}$ hygromycin B and 400 $\mu\text{g/ml}$ G418. The doubling time of both cell lines was about 24-40 hours, typically about 32 hours.

- Typically, cell monolayers are passaged when they reach about 60-
- 20 80% confluence and are split every 48-72 hours. Cells that reach greater than 80% confluence senesce in culture and are not preferred for chromosome harvesting. Cells may be plated in 100-200 100-mm dishes at about 50-70% confluency 12-30 hours before mitotic arrest (see, below).

- Other cell lines that may be used as hosts for artificial chromo-
- 25 somes and from which the artificial chromosomes may be isolated include, but are not limited to, PtK1 (NBL-3) marsupial kidney cells (ATCC accession no. CCL35), CHO-K1 Chinese hamster ovary cells (ATCC accession no. CCL61), V79-4 Chinese hamster lung cells (ATCC accession no. CCL93), Indian muntjac skin cells (ATCC accession no. CCL157),
- 30 LMTK(-) thymidine kinase deficient murine L cells (ATCC accession no. CCL1.3), Sf9 fall armyworm (*Spodoptera frugiperda*) ovary cells (ATCC

accession no. CRL 1711) and any generated heterokaryon (hybrid) cell lines, such as, for example, the hamster-murine hybrid cells described herein, that may be used to construct MACs, particularly SATACs.

- Cell lines may be selected, for example, to enhance efficiency of
- 5 artificial chromosome production and isolation as may be desired in large-scale production processes. For instance, one consideration in selecting host cells may be the artificial chromosome-to-total chromosome ratio of the cells. To facilitate separation of artificial chromosomes from endogenous chromosomes, a higher artificial chromosome-to-total
- 10 chromosome ratio might be desirable. For example, for H1D3 cells (a murine/hamster heterokaryon; see Figure 4), this ratio is 1:50, i.e., one artificial chromosome (the megachromosome) to 50 total chromosomes. In contrast, Indian muntjac skin cells (ATCC accession no. CCL157)
- 15 contain a smaller total number of chromosomes (a diploid number of chromosomes of 7), as do kangaroo rat cells (a diploid number of chromosomes of 12) which would provide for a higher artificial chromosome-to-total chromosome ratio upon introduction of, or generation of, artificial chromosomes in the cells.

- Another consideration in selecting host cells for production and
- 20 isolation of artificial chromosomes may be size of the endogenous chromosomes as compared to that of the artificial chromosomes. Size differences of the chromosomes may be exploited to facilitate separation of artificial chromosomes from endogenous chromosomes. For example, because Indian muntjac skin cell chromosomes are considerably larger
- 25 than minichromosomes and truncated megachromosomes, separation of the artificial chromosome from the muntjac chromosomes may possibly be accomplished using univariate (one dye, either Hoechst 33258 or Chromomycin A3) FACS separation procedures.

- Another consideration in selecting host cells for production and iso-
- 30 lation of artificial chromosomes may be the doubling time of the cells. For example, the amount of time required to generate a sufficient number

of artificial chromosome-containing cells for use in procedures to isolate artificial chromosomes may be of significance for large-scale production. Thus, host cells with shorter doubling times may be desirable. For instance, the doubling time of V79 hamster lung cells is about 9-10 hours in comparison to the approximately 32-hour doubling time of H1D3 cells.

Accordingly, several considerations may go into the selection of host cells for the production and isolation of artificial chromosomes. It may be that the host cell selected as the most desirable for de novo formation of artificial chromosomes is not optimized for large-scale production of the artificial chromosomes generated in the cell line. In such cases, it may be possible, once the artificial chromosome has been generated in the initial host cell line, to transfer it to a production cell line more well suited to efficient, high-level production and isolation of the artificial chromosome. Such transfer may be accomplished through several methods, for example through microcell fusion, as described herein, or microinjection into the production cell line of artificial chromosomes purified from the generating cell line using procedures such as described herein. Production cell lines preferably contain two or more copies of the artificial artificial chromosome per cell.

20 B. Chromosome isolation

In general, cells are typically cultured for two generations at exponential growth prior to mitotic arrest. To accumulate mitotic 1B3 and GHB42 cells in one particular isolation procedure, 5 μ g/ml colchicine was added for 12 hours to the cultures. The mitotic index obtained was 60-80%. The mitotic cells were harvested by selective detachment by gentle pipetting of the medium on the monolayer cells. It is also possible to utilize mechanical shake-off as a means of releasing the rounded-up

(mitotic) cells from the plate. The cells were sedimented by centrifugation at $200 \times g$ for 10 minutes.

- Cells (grown on plastic or in suspension) may be arrested in different stages of the cell cycle with chemical agents other than
- 5 colchicine, such as hydroxyurea, vinblastine, colcemid or aphidicolin. Chemical agents that arrest the cells in stages other than mitosis, such as hydroxyurea and aphidicolin, are used to synchronize the cycles of all cells in the population and then are removed from the cell medium to allow the cells to proceed, more or less simultaneously, to mitosis at
 - 10 which time they may be harvested to disperse the chromosomes. Mitotic cells could be enriched for a mechanical shake-off (adherent cells). The cell cycles of cells within a population of MAC-containing cells may also be synchronized by nutrient, growth factor or hormone deprivation which leads to an accumulation of cells in the G_1 or G_0 stage; readdition of
 - 15 nutrients or growth factors then allows the quiescent cells to re-enter the the cell cycle in synchrony for about one generation. Cell lines that are known to respond to hormone deprivation in this manner, and which are suitable as hosts for artificial chromosomes, include the Nb2 rat lymphoma cell line which is absolutely dependent on prolactin for
 - 20 stimulation of proliferation (see Gout et al. (1980) Cancer Res. 40:2433-2436). Culturing the cells in prolactin-deficient medium for 18-24 hours leads to arrest of proliferation, with cells accumulating early in the G_1 phase of the cell cycle. Upon addition of prolactin, all the cells progress through the cell cycle until M phase at which point greater than 90% of
 - 25 the cells would be in mitosis (addition of colchicine could increase the amount of the mitotic cells to greater than 95%). The time between reestablishing proliferation by prolactin addition and harvesting mitotic cells for chromosome separation may be empirically determined.

- Alternatively, adherent cells, such as V79 cells, may be grown in
- 30 roller bottles and mitotic cells released from the plastic surface by rotating the roller bottles at 200 rpm or greater (Shwarchuk et al. (1993) Int. J.

Radiat. Biol. 64:601-612). At any given time, approximately 1% of the cells in an exponentially growing asynchronous population is in M-phase. Even without the addition of colchicine, 2×10^7 mitotic cells have been harvested from four 1750-cm² roller bottles after a 5-min spin at 200 rpm. Addition of colchicine for 2 hours may increase the yield to 6×10^8 mitotic cells.

- Several procedures may be used to isolate metaphase chromosomes from these cells, including, but not limited to, one based on a polyamine buffer system [Cram et al. (1990) Methods in Cell Biology 33:377-382], one on a modified hexylene glycol buffer system [Hadlaczky et al. (1982) Chromosoma 86:643-65], one on a magnesium sulfate buffer system [Van den Engh et al. (1988) Cytometry 9:266-270 and Van den Engh et al. (1984) Cytometry 5:108], one on an acetic acid fixation buffer system [Stoeht et al. (1982) Histochemistry 74:57-61], and one on a technique utilizing hypotonic KCl and propidium iodide [Cram et al. (1994) XVII meeting of the International Society for Analytical Cytology, October 16-21, Tutorial IV Chromosome Analysis and Sorting with Commerical Flow Cytometers; Cram et al. (1990) Methods in Cell Biology 33:376].

20 1. Polyamine procedure

- In the polyamine procedure that was used in isolating artificial chromosomes from either 1B3 or GHB42 cells, about 10^7 mitotic cells were incubated in 10 ml hypotonic buffer (75 mM KCl, 0.2 mM spermine, 0.5 mM spermidine) for 10 minutes at room temperature to swell the cells. The cells are swollen in hypotonic buffer to loosen the metaphase chromosomes but not to the point of cell lysis. The cells were then centrifuged at 100 x g for 8 minutes, typically at room temperature. The cell pellet was drained carefully and about 10^7 cells were resuspended in 1 ml polyamine buffer [15 mM Tris-HCl, 20 mM NaCl, 80 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 14 mM β -mercaptoethanol, 0.1% digitonin, 0.2 mM Spermine, 0.5 mM spermidine] for physical dispersal of the

metaphase chromosomes. Chromosomes were then released by gently drawing the cell suspension up and expelling it through a 22 G needle attached to a 3 ml plastic syringe. The chromosome concentration was about $1-3 \times 10^8$ chromosomes/ml.

- 5 The polyamine buffer isolation protocol is well suited for obtaining high molecular weight chromosomal DNA [Sillar and Young (1981) J. Histochem. Cytochem. 29:74-78; VanDilla et al. (1986) Biotechnology 4:537-552; Bartholdi et al. (1988) In "Molecular Genetics of Mammalian Cells" (M.Goettsman, ed.), Methods in Enzymology 151:252-267.
- 10 Academic Press, Orlando]. The chromosome stabilizing buffer uses the polyamines spermine and spermidine to stabilize chromosome structure [Blumenthal et al. (1979)J. Cell Biol. 81:255-259; Lalande et al. (1985) Cancer Genet. Cytogenet. 23:151-157] and heavy metals chelators to reduce nuclease activity.
- 15 The polyamine buffer protocol has wide applicability, however, as with other protocols, the following variables must be optimized for each cell type: blocking time, cell concentration, type of hypotonic swelling buffer, swelling time, volume of hypotonic buffer, and vortexing time. Chromosomes prepared using this protocol are typically highly condensed.
- 20 There are several hypotonic buffers that may be used to swell the cells, for example buffers such as the following: 75 mM KCl; 75 mM KCl, 0.2 mM spermine, 0.5 mM spermidine; Ohnuki's buffer of 16.2 mM sodium nitrate, 6.5 mM sodium acetate, 32.4 mM KCl [Ohnuki (1965) Nature 208:916-917 and Ohnuki (1968) Chromosoma 25:402-428]; and
- 25 a variation of Ohnuki's buffer that additionally contains 0.2 mM spermine and 0.5 mM spermidine. The amount and hypotonicity of added buffer vary depending on cell type and cell concentration. Amounts may range from 2.5 - 5.5 ml per 10^7 cells or more. Swelling times may vary from 10-90 minutes depending on cell type and which swelling buffer is used.
- 30 The composition of the polyamine isolation buffer may also be varied. For example, one modified buffer contains 15 mM Tris-HCl, pH

7.2, 70 mM NaCl, 80 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 14 mM beta-mercaptoethanol, 0.25% Triton-X, 0.2 mM spermine and 0.5 mM spermidine.

- Chromosomal dispersal may also be accomplished by a variety of
- 5 physical means. For example, cell suspension may be gently drawn up and expelled in a 3-ml syringe fitted with a 22-gauge needle [Cram et al. (1990) Methods in Cell Biology 33:377-382], cell suspension may be agitated on a bench-top vortex [Cram et al. (1990) Methods in Cell Biology 33:377-382], cell suspension may be disrupted with a
- 10 homogenizer [Sillar and Young (1981) J. Histochem. Cytochem. 29:74-78; Carrano et al. (1979) Proc. Natl. Acad. Sci. U.S.A. 76:1382-1384] and cell suspension may be disrupted with a bench-top ultrasonic bath [Stoehr et al. (1982) Histochemistry 74:57-61].

2. Hexylene glycol buffer system

- 15 In the hexylene glycol buffer procedure that was used in isolating artificial chromosomes from either 1B3 or GHB42 cells, about 8×10^8 mitotic cells were resuspended in 10 ml glycine-hexylene glycol buffer [100 mM glycine, 1% hexylene glycol, pH 8.4-8.6 adjusted with saturated Ca-hydroxide solution] and incubated for 10 minutes at 37°C,
- 20 followed by centrifugation for 10 minutes to pellet the nuclei. The supernatant was centrifuged again at 200 x g for 20 minutes to pellet the chromosomes. Chromosomes were resuspended in isolation buffer ($1-3 \times 10^8$ chromosomes/ml).

- The hexylene glycol buffer composition may also be modified. For
- 25 example, one modified buffer contains 25 mM Tris-HCl, pH 7.2, 750 mM hexylene glycol, 0.5 mM CaCl_2 , 1.0 mM MgCl_2 [Carrano et al. (1979) Proc. Natl. Acad. Sci. U.S.A. 76:1382-1384].

3. Magnesium-sulfate buffer system

- This buffer system may be used with any of the methods of cell
- 30 swelling and chromosomal dispersal, such as described above in connection with the polyamine and hexylene glycol buffer systems. In

this procedure, mitotic cells are resuspended in the following buffer: 4.8 mM HEPES, pH 8.0, 9.8 mM MgSO_4 , 48 mM KCl, 2.9 mM dithiothreitol [Van den Engh *et al.* (1985) *Cytometry* 6:92 and Van den Engh *et al.* (1984) *Cytometry* 5:108].

5 4. Acetic acid fixation buffer system

This buffer system may be used with any of the methods of cell swelling and chromosomal dispersal, such as described above in connection with the polyamine and hexylene glycol buffer systems. In this procedure, mitotic cells are resuspended in the following buffer: 25
10 mM Tris-HCl, pH 3.2, 750 mM (1,6)-hexandiol, 0.5 mM CaCl_2 , 1.0% acetic acid [Stoehr *et al.* (1982) *Histochemistry* 74:57-61].

5. KCl-propidium iodide buffer system

This buffer system may be used with any of the methods of cell swelling and chromosomal dispersal, such as described above in
15 connection with the polyamine and hexylene glycol buffer systems. In this procedure, mitotic cells are resuspended in the following buffer: 25 mM KCl, 50 $\mu\text{g/ml}$ propidium iodide, 0.33% Triton X-100, 333 $\mu\text{g/ml}$ RNase [Cram *et al.* (1990) *Methods in Cell Biology* 33:376].

The fluorescent dye propidium iodide is used and also serves as a
20 chromosome stabilizing agent. Swelling of the cells in the hypotonic medium (which may also contain propidium iodide) may be monitored by placing a small drop of the suspension on a microscope slide and observing the cells by phase/fluorescent microscopy. The cells should exclude the propidium iodide while swelling, but some may lyse
25 prematurely and show chromosome fluorescence. After the cells have been centrifuged and resuspended in the KCl-propidium iodide buffer system, they will be lysed due to the presence of the detergent in the buffer. The chromosomes may then be dispersed and then incubated at 37°C for up to 30 minutes to permit the RNase to act. The chromosome
30 preparation is then analyzed by flow cytometry. The propidium iodide fluorescence can be excited at the 488 nm wavelength of an argon laser

and detected through an OG 570 optical filter by a single photomultiplier tube. The single pulse may be integrated and acquired in an univariate histogram. The flow cytometer may be aligned to a CV of 2% or less using small (1.5 μm diameter) microspheres. The chromosome preparation

5 is filtered through 60 μm nylon mesh before analysis.

C. Staining of chromosomes with DNA-specific dyes

Subsequent to isolation, the chromosome preparation was stained with Hoechst 33258 at 6 $\mu\text{g}/\text{ml}$ and chromomycin A3 at 200 $\mu\text{g}/\text{ml}$.

Fifteen minutes prior to analysis, 25 mM Na-sulphite and 10 mM Na-
10 citrate were added to the chromosome suspension.

D. Flow sorting of chromosomes

Chromosomes obtained from 1B3 and GHB42 cells and maintained were suspended in a polyamine-based sheath buffer (0.5 mM EGTA, 2.0 mM EDTA, 80 mM KCl, 70 mM NaCl, 15 mM Tris-HCl, pH 7.2, 0.2 mM

15 spermine and 0.5 mM spermidine) [Sillar and Young (1981) J. Histochem. Cytochem. 29:74-78]. The chromosomes were then passed through a dual-laser cell sorter [FACStar Plus or FAXStar Vantage Becton Dickinson Immunocytometry System; other dual-laser sorters may also be used, such as those manufactured by Coulter Electronics (Elite ESP) and
20 Cytomation (MoFlo)] in which two lasers were set to excite the dyes separately, allowing a bivariate analysis of the chromosome by size and base-pair composition. Because of the difference between the base composition of the SATACs and the other chromosomes and the resulting difference in interaction with the dyes, as well as size differences, the

25 SATACs were separated from the other chromosomes.

E. Storage of the sorted artificial chromosomes

Sorted chromosomes may be pelleted by centrifugation and resuspended in a variety of buffers, and stored at 4°C. For example, the isolated artificial chromosomes may be stored in GH buffer (100 mM

30 glycine, 1% hexylene glycol pH 8.4—8.6 adjusted with saturated Ca-hydroxide solution) [see, e.g., Hadlaczy et al. (1982) Chromosoma

86:643-659] for one day and embedded by centrifugation into agarose.

The sorted chromosomes were centrifuged into an agarose bed and the plugs are stored in 500 mM EDTA at 4° C. Additional storage buffers include CMB-I/polyamine buffer (17.5 mM Tris-HCl, pH 7.4, 1.1 mM

- 5 EDTA, 50 mM epsilon-amino caproic acid, 5 mM benzamide-HCl, 0.40 mM spermine, 1.0 mM spermidine, 0.25 mM EGTA, 40 mM KCl, 35 mM NaCl) and CMB-II/polyamine buffer (100 mM glycine, pH 7.5, 78 mM hexylene glycol, 0.1 mM EDTA, 50 mM epsilon-amino caproic acid, 5 mM benzamide-HCl, 0.40 mM spermine, 1.0 mM spermidine, 0.25 mM EGTA, 10 40 mM KCl, 35 mM NaCl).

When microinjection is the intended use, the sorted chromosomes are stored in 30% glycerol at -20° C. Sorted chromosomes may also be stored without glycerol for short periods of time (3-6 days) in storage buffers at 4°C. Exemplary buffers for microinjection include CBM-I (10

- 15 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 50 mM epsilon-amino caproic acid, 5 mM benzamide-HCl, 0.30 mM spermine, 0.75 mM spermidine), CBM-II (100 mM glycine, pH 7.5, 78 mM hexylene glycol, 0.1 mM EDTA, 50 mM epsilon-amino caproic acid, 5 mM benzamide-HCl, 0.30 mM spermine, 0.75 mM spermidine).

- 20 For long-term storage of sorted chromosomes, the above buffers are preferably supplemented with 50% glycerol and stored at -20°C.

F. Quality control

1. Analysis of the purity

The purity of the sorted chromosomes was checked by

- 25 fluorescence *in situ* hybridization (FISH) with a biotin-labeled mouse satellite DNA probe [see, Hadlaczky *et al.* (1991) Proc. Natl. Acad. Sci. U.S.A. 88:8106-8110]. Purity of the isolated chromosomes was about 97—99%.

2. Characteristics of the sorted chromosomes

Pulsed field gel electrophoresis and Southern hybridization were carried out to determine the size distribution of the DNA content of the sorted artificial chromosomes.

5 G. Functioning of the purified artificial chromosomes

To check whether their activity is preserved, the purified artificial chromosomes may be microinjected (using methods such as those described in Example 13) into primary cells, somatic cells and stem cells which are then analyzed for expression of the heterologous genes carried
10 by the artificial chromosomes, e.g., such as analysis for growth on selective medium and assays of β -galactosidase activity.

II. Sorting of mammalian artificial chromosome-containing microcells

A. Micronucleation

Cells were grown to 80–90% confluency in 4 T150 flasks.

- 15 Colcemid was added to a final concentration of 0.06 $\mu\text{g/ml}$, and then incubated with the cells at 37°C for 24 hours.

B. Enucleation

Ten $\mu\text{g/ml}$ cytochalasin B was added and the resulting microcells were centrifuged at 15,000 rpm for 70 minutes at 28–33° C.

20 C. Purification of microcells by filtration

The microcells were purified using Swinnex filter units and Nucleopore filters [5 μm and 3 μm].

D. Staining and sorting microcells

- As above, the cells were stained with Hoechst and chromomycin
25 A3 dyes. The microcells were sorted by cell sorter to isolate the microcells that contain the mammalian artificial chromosomes.

E. Fusion

- The microcells that contain the artificial chromosome are fused, for example, as described in Example 1.A.5., to selected primary cells,
30 somatic cells, embryonic stem cells to generate transgenic (non-human)

animals and for gene therapy purposes, and to other cells to deliver the chromosomes to the cells.

EXAMPLE 11

Introduction of mammalian artificial chromosomes into insect cells

- 5 Insect cells are useful hosts for MACs, particularly for use in the production of gene products, for a number of reasons, including:
1. A mammalian artificial chromosome provides an extra-genomic specific integration site for introduction of genes encoding proteins of interest [reduced chance of mutation in production system].
 - 10 2. The large size of an artificial chromosome permits megabase size DNA integration so that genes encoding an entire pathway leading to a protein or nonprotein of therapeutic value, such as an alkaloid [digitalis, morphine, taxol] can be accomodated by the artificial chromosome.
 - 15 3. Amplification of genes encoding useful proteins can be accomplished in the artificial mammalian chromosome to obtain higher protein yields in insect cells.
 4. Insect cells support required post-translational modifications (glycosylation, phosphorylation) essential for protein biological function.
 5. Insect cells do not support mammalian viruses — eliminates
20 cross-contamination of product with human infectious agents.
 6. The ability to introduce chromosomes circumvents traditional recombinant baculovirus systems for production of nutritional, industrial or medicinal proteins in insect cell systems.
 7. The low temperature optimum for insect cell growth (28° C)
25 permits reduced energy cost of production.
 8. Serum free growth medium for insect cells will result in lower production costs.
 9. Artificial chromosome-containing cells can be stored indefinitely at low temperature.

10. Insect larvae will serve as biological factories for the production of nutritional, medicinal or industrial proteins by microinjection of fertilized insect eggs.

A. Demonstration that insect cells recognize mammalian promoters

- 5 Gene constructs containing a mammalian promoter, such as the CMV promoter, linked to a detectable marker gene [*Renilla* luciferase gene (see, e.g., U.S. Patent No. 5,292,658 for a description of DNA encoding the *Renilla* luciferase, and plasmid pTZrLuc-1, which can provide the starting material for construction of such vectors, see also SEQ ID No. 10] and also including the simian virus 40 (SV40) promoter operably linked to the β -galactosidase gene were introduced into the cells of two species *Trichoplusia ni* [cabbage looper] and *Bombyx mori* [silkworm].

- After transferring the constructs into the insect cell lines either by electroporation or by microinjection, expression of the marker genes was 15 detected in luciferase assays (see e.g., Example 12.C.3) and in β -galactosidase assays (such as lacZ staining assays) after a 24-h incubation. In each case a positive result was obtained in the samples containing the genes which was absent in samples in which the genes were omitted. In addition, a *B. mori* β -actin promoter-*Renilla* luciferase 20 gene fusion was introduced into the *T. ni* and *B. mori* cells which yielded light emission after transfection. Thus, certain mammalian promoters function to direct expression of these marker genes in insect cells. Therefore, MACs are candidates for expression of heterologous genes in insect cells.

25 B. Construction of vectors for use in insect cells and fusion with mammalian cells

1. Transform LMTK⁻ cells with expression vector with:
 - a. *B. mori* β -actin promoter— Hyg^r selectable marker gene for insect cells, and
 - 30 b. SV40 or CMV promoters controlling a puromycin^r selectable marker gene for mammalian cells.

2. Detect expression of the mammalian promoter in LMTK cells (puromycin^r LMTK cells)
3. Use puromycin^r cells in fusion experiments with *Bombyx* and *Trichoplusia* cells, select Hyg^r cells.

5 C. Insertion of the MACs into insect cells

These experiments are designed to detect expression of a detectable marker gene [such as the β -galactosidase gene expressed under the control of a mammalian promoter, such as pSV40] located on a MAC that has been introduced into an insect cell. Data indicate that β -gal was expressed.

- 10 Insect cells are fused with mammalian cells containing mammalian artificial chromosomes, e.g., the minichromosome [EC3/7C5] or the mini and the megachromosome [such as GHB42, which is a cell line recloned from G3D5] or a cell line that carries only the megachromosome [such as
- 15 H1D3 or a reclone therefrom]. Fusion is carried out as follows:

1. mammalian + insect cells (50/50%) in log phase growth are mixed;
2. calcium/PEG cell fusion: (10 min — 0.5 h);
3. heterokaryons (+ 72 h) are selected.

- 20 The following selection conditions to select for insect cells that contain a MAC can be used: [+ = positive selection; - = negative selection]:

1. growth at 28° C (+ insect cells, - mammalian cells);
2. Graces insect cell medium [SIGMA] (- mammalian cells);
- 25 3. no exogenous CO₂ (- mammalian cells); and/or
4. antibiotic selection (Hyg or G418) (+ transformed insect cells).

- Immediately following the fusion protocol, many heterokaryons [fusion events] are observed between the mammalian and each species of insect cells [up to 90% heterokaryons]. After growth [2+ weeks] on
- 30 insect medium containing G418 and/or hygromycin at selection levels used for selection of transformed mammalian cells, individual colonies are

detected growing on the fusion plates. By virtue of selection for the antibiotic resistance conferred by the MAC and selection for insect cells, these colonies should contain MACs.

- The *B. mori* β -actin gene promoter has been shown to direct
- 5 expression of the β -galactosidase gene in *B. mori* cells and mammalian cells (e.g., EC3/7C5 cells). The *B. mori* β -actin gene promoter is, thus, particularly useful for inclusion in MACs generated in mammalian cells that will subsequently be transferred into insect cells because the presence of any marker gene linked to the promoter can be determined in
- 10 the mammalian and resulting insect cell lines.

EXAMPLE 12

Preparation of chromosome fragmentation vectors and other vectors for targeted integration of DNA into MACs

- Fragmentation of the megachromosome should ultimately result in
- 15 smaller stable chromosomes that contain about 15 Mb to 50 Mb that will be easily manipulated for use as vectors. Vectors to effect such fragmentation should also aid in determination and identification of the elements required for preparation of an in vitro-produced artificial chromosome.
- 20 Reduction in the size of the megachromosome can be achieved in a number of different ways including: stress treatment, such as by starvation, or cold or heat treatment; treatment with agents that destabilize the genome or nick DNA, such as BrdU, coumarin, EMS and others; treatment with ionizing radiation [see, e.g., Brown (1992) Curr.
- 25 Opin. Genes Dev. 2:479-486]; and telomere-directed in vivo chromosome fragmentation [see, e.g., Farr et al. (1995) EMBO J. 14:5444-5454].

A. Preparation of vectors for fragmentation of the artificial chromosome and also for targeted integration of selected gene products

30 1. Construction of pTEMPUD

Plasmid pTEMPUD [see Figure 5] is a mouse homologous recombination "killer" vector for in vivo chromosome fragmentation, and

also for inducing large-scale amplification via site-specific integration. With reference to Figure 5, the ~3,625-bp Sall-PstI fragment was derived from the pBabe-puro retroviral vector [see, Morgenstern et al. (1990) Nucleic Acids Res. 18:3587-3596]. This fragment contains DNA

- 5 encoding ampicillin resistance, the pUC origin of replication, and the puromycin N-acetyl transferase gene under control of the SV40 early promoter. The URA3 gene portion comes from the pYAC5 cloning vector [SIGMA]. URA3 was cut out of pYAC5 with Sall-XhoI digestion, cloned into pNEB193 [New England Biolabs], which was then cut with EcoRI-Sall
10 and ligated to the Sall site of pBabepuro to produce pPU.

A 1293-bp fragment [see SEQ ID No. 1] encoding the mouse major satellite, was isolated as an EcoRI fragment from a DNA library produced from mouse LMTK⁺ fibroblast cells and inserted into the EcoRI site of pPU to produce pMPU.

- 15 The TK promoter-driven diphtheria toxin gene [DT-A] was derived from pMC1DT-A [see, Maxwell et al. (1986) Cancer Res. 46:4660-4666] by BglII-XhoI digestion and cloned into the pMC1neo poly A expression vector [STRATAGENE, La Jolla, CA] by replacing the neomycin-resistance gene coding sequence. The TK promoter, DT-A gene and poly A
20 sequence were removed from this vector, cohesive ends were filled with Klenow and the resulting fragment blunt end-ligated and ligated into the SnaBI (TACGTA) of pMPU to produce pMPUD.

- The Htel 2.5-kb fragment [see SEQ ID No.3] was inserted at the PstI site [see the 6100 PstI - 3625 PstI fragment on pTEMPUD] of
25 pMPUD to produce pTEMPUD. This fragment includes a human telomere. It includes a unique BglII site [see nucleotides 1042-1047 of SEQ ID No.3], which will be used as a site for introduction of a synthetic telomere that includes multiple repeats [80] of TTAGGG with BamHI and BglII ends for insertion into the BglII site which will then remain unique,
30 since the BamHI overhang is compatible with the BglII site. Ligation of a BamHI fragment to a BglII destroys the BglII site, so that only a single

BglIII site will remain. Selection for the unique BglIII site insures that the synthetic telomere will be inserted in the correct orientation. The unique BglIII site is the site at which the vector is linearized.

- To generate a synthetic telomere made up of multiple repeats of
- 5 the sequence TTAGGG, attempts were made to clone or amplify ligation products of 30-mer oligonucleotides containing repeats of the sequence. Two 30-mer oligonucleotides, one containing four repeats of TTAGGG bounded on each end of the complete run of repeats by half of a repeat and the other containing five repeats of the complement AATCCC, were
- 10 annealed. The resulting double-stranded molecule with 3-bp protruding ends, each representing half of a repeat, was expected to ligate with itself to yield concatamers of $n \times 30$ bp. However, this approach was unsuccessful, likely due to formation of quadruplex DNA from the G-rich strand. Similar difficulty has been encountered in attempts to generate
- 15 long repeats of the pentameric human satellite II and III units. Thus, it appears that, in general, any oligomer sequence containing periodically spaced consecutive series of guanine nucleotides is likely to form undesired quadruplex formation that hinders construction of long double-stranded DNAs containing the sequence.
- 20 Therefore, in another attempt to construct a synthetic telomere for insertion into the BglIII site of pTEMPUD, the starting material was based on the complementary C-rich repeat sequence (i.e., AATCCC) which would not be susceptible to quadruplex structure formation. Two plasmids, designated pTEL280110 and pTel280111, were constructed as
- 25 follows to serve as the starting materials.

- First, a long oligonucleotide containing 9 repeats of the sequence AATCCC (i.e., the complement of telomere sequence TTAGGG) in reverse order bounded on each end of the complete run of repeats by half of a repeat (therefore, in essence, containing 10 repeats), and recognition
- 30 sites for PstI and PacI restriction enzymes was synthesized using standard methods. The oligonucleotide sequence is as follows:

5'-AAACTGCAGGTTAATTAACCCTAACCCCTAACCCCTAACCCCTAAC
CCTAACCCCTAACCCCTAACCCCTAACCCGGGAT-3' (SEQ ID NO. 29)

A partially complementary short oligonucleotide of sequence

3'-TTGGGCCCTAGGCTTAAGG-5' (SEQ ID NO. 30)

- 5 was also synthesized. The oligonucleotides were gel-purified, annealed, repaired with Klenow polymerase and digested with EcoRI and PstI. The resulting EcoRI/PstI fragment was ligated with EcoRI/PstI-digested pUC19. The resulting plasmid was used to transform E. coli DH5 α competent cells and plasmid DNA (pTel102) from one of the transformants surviving
- 10 selection on LB/ampicillin was digested with PacI, rendered blunt-ended by Klenow and dNTPs and digested with HindIII. The resulting 2.7-kb fragment was gel-purified.

- Simultaneously, the same plasmid was amplified by the polymerase chain reaction using extended and more distal 26-mer M13
- 15 sequencing primers. The amplification product was digested with SmaI and HindIII, the double-stranded 84-bp fragment containing the 60-bp telomeric repeat (plus 24 bp of linker sequence) was isolated on a 6% native polyacrylamide gel, and ligated with the double-digested pTel102 to yield a 120-bp telomeric sequence. This plasmid was used to
- 20 transform DH5 α cells. Plasmid DNA from two of the resulting recombinants that survived selection on ampicillin (100 μ g/ml) was sequenced on an ABI DNA sequencer using the dye-termination method. One of the plasmids, designated pTel29, contained a sequence of 20
- 25 repeats of the sequence TTAGGG (i.e., 19 successive repeats of TTAGGG bounded on each end of the complete run of repeats with half of a repeat). The other plasmid, designated pTel28, had undergone a deletion of 2 bp (TA) at the junction where the two sequences, each containing, in essence, 10 repeats of the TTAGGG sequence, that had been ligated to yield the plasmid. This resulted in a GGGTGGG motif at the junction in
- 30 pTel28. This mutation provides a useful tag in telomere-directed chromosome fragmentation experiments. Therefore, the pTel29 insert

was amplified by PCR using pUC/M13 sequencing primers based on sequence somewhat longer and farther from the polylinker than usual as follows:

5'-GCCAGGGTTTTCCCAGTCACGACGT-3' (SEQ ID NO. 31)

- 5 or in some experiments

5'-GCTGCAAGGCGATTAAGTTGGGTAAC-3' (SEQ ID NO. 32)

as the m13 forward primer, and

5'-TATGTTGTGTGGAATTGTGAGCGGAT-3' (SEQ ID NO. 33)

as the m13 reverse primer.

- 10 The amplification product was digested with SmaI and HindIII. The resulting 144-bp fragment was gel-purified on a 6% native polyacrylamide gel and ligated with pTel28 that had been digested with PacI, blunt-ended with Klenow and dNTP and then digested with HindIII to remove linker. The ligation yielded a plasmid designated pTel2801 containing a telomeric
- 15 sequence of 40 repeats of the sequence TTAGGG in which one of the repeats (i.e., the 30th repeat) lacked two nucleotides (TA), due to the deletion that had occurred in pTel28, to yield a repeat as follows: TGGG.

- In the next extension step, pTel2801 was digested with SmaI and HindIII and the 264-bp insert fragment was gel-purified and ligated with
- 20 pTel2801 which had been digested with PacI, blunt-ended and digested with HindIII. The resulting plasmid was transformed into DH5 α cells and plasmid DNA from 12 of the resulting transformants that survived selection on ampicillin was examined by restriction enzyme analysis for the presence of a 0.5-kb EcoRI/PstI insert fragment. Eleven of the
- 25 recombinants contained the expected 0.5-kb insert. The inserts of two of the recombinants were sequenced and found to be as expected. These plasmids were designated pTel280110 and pTel280111. These plasmids, which are identical, both contain 80 repeats of the sequence TTAGGG, in which two of the repeats (i.e., the 30th and 70th repeats) lacked two
- 30 nucleotides (TA), due to the deletion that had occurred in pTel28, to yield a repeat as follows: TGGG. Thus, in each of the cloning steps (except

- the first), the length of the synthetic telomere doubled; that is, it was increasing in size exponentially. Its length was 60×2^n bp, wherein n is the number of extension cloning steps undertaken. Therefore, in principle (assuming *E. coli*, or any other microbial host, e.g., yeast, tolerates long tandem repetitive DNA), it is possible to assemble any desirable size of safe telomeric repeats.

- In a further extension step, pTel280110 was digested with PacI, blunt-ended with Klenow polymerase in the presence of dNTP, then digested with HindIII. The resulting 0.5-kb fragment was gel purified.
- 10 Plasmid pTel280111 was cleaved with SmaI and HindIII and the 3.2-kb fragment was gel-purified and ligated to the 0.5-kb fragment from pTel280110. The resulting plasmid was used to transform DH5 α cells. Plasmid DNA was purified from transformants surviving ampicillin selection. Nine of the selected recombinants were examined by
- 15 restriction enzyme analysis for the presence of a 1.0-kb EcoRI/PstI fragment. Four of the recombinants (designated pTik2, pTik6, pTik7 and pTik8) were thus found to contain the desired 960 bp telomere DNA insert sequence that included 160 repeats of the sequence TTAGGG in which four of the repeats lacked two nucleotides (TA), due to the deletion
- 20 that had occurred in pTel28, to yield a repeat as follows: TGGG. Partial DNA sequence analysis of the EcoRI/PstI fragment of two of these plasmids (i.e., pTik2 and pTik6), in which approximately 300 bp from both ends of the fragment were elucidated, confirmed that the sequence was composed of successive repeats of the TTAGGG sequence.
- 25 In order to add PmeI and BglII sites to the synthetic telomere sequence, pTik2 was digested with PacI and PstI and the 3.7-kb fragment (i.e., 2.7-kb pUC19 and 1.0-kb repeat sequence) was gel-purified and ligated at the PstI cohesive end with the following oligonucleotide 5'-GGGTTTAAACAGATCTCTGCA-3' (SEQ ID NO. 34). The ligation product
- 30 was subsequently repaired with Klenow polymerase and dNTP, ligated to itself and transformed into *E. coli* strain DH5 α . A total of 14

recombinants surviving selection on ampicillin were obtained. Plasmid DNA from each recombinant was able to be cleaved with BglIII indicating that this added unique restriction site had been retained by each recombinant. Four of the 14 recombinants contained the complete 1-kb

- 5 synthetic telomere insert, whereas the insert of the remaining 10 recombinants had undergone deletions of various lengths. The four plasmids in which the 1-kb synthetic telomere sequence remained intact were designated pTikV2, pTikV5, pTikV8 and pTikV12. Each of these plasmids could also be digested with PmeI; in addition the presence of
- 10 both the BglIII and PmeI sites was verified by sequence analysis. Any of these four plasmids can be digested with BamHI and BglIII to release a fragment containing the 1-kb synthetic telomere sequence which is then ligated with BglIII-digested pTEMPUD.

2. Use of pTEMPUD for in vivo chromosome fragmentation

- 15 Linearization of pTEMPUD by BglIII results in a linear molecule with a human telomere at one end. Integration of this linear fragment into the chromosome, such as the megachromosome in hybrid cells or any mouse chromosome which contains repeats of the mouse major satellite sequence results in integration of the selectable marker puromycin-
- 20 resistance gene and cleavage of the plasmid by virtue of the telomeric end. The DT gene prevents that entire linear fragment from integrating by random events, since upon integration and expression it is toxic. Thus random integration will be toxic, so site-directed integration into the targeted DNA will be selected. Such integration will produce fragmented
- 25 chromosomes.

The fragmented truncated chromosome with the new telomere will survive, and the other fragment without the centromere will be lost.

Repeated in vivo fragmentations will ultimately result in selection of the smallest functioning artificial chromosome possible. Thus, this vector

- 30 can be used to produce minichromosomes from mouse chromosomes, or to fragment the megachromosome. In principle, this vector can be used

to target any selected DNA sequence in any chromosome to achieve fragmentation.

3. Construction of pTERPUD

A fragmentation/targeting vector analogous to pTEMPUD for in vivo chromosome fragmentation, and also for inducing large-scale amplification via site-specific integration but which is based on mouse rDNA sequence instead of mouse major satellite DNA has been designated pTERPUD. In this vector, the mouse major satellite DNA sequence of pTEMPUD has been replaced with a 4770-bp BamHI fragment of megachromosome clone 161 which contains sequence corresponding to nucleotides 10,232-15,000 in SEQ ID NO. 16.

4. pHASPUD and pTEMPHu3

Vectors that specifically target human chromosomes can be constructed from pTEMPUD. These vectors can be used to fragment specific human chromosomes, depending upon the selected satellite sequence, to produce human minichromosomes, and also to isolate human centromeres.

a. pHASPUD

To render pTEMPUD suitable for fragmenting human chromosomes, the mouse major satellite sequence is replaced with human satellite sequences. Unlike mouse chromosomes, each human chromosome has a unique satellite sequence. For example, the mouse major satellite has been replaced with a human hexameric α -satellite [or alphoid satellite] DNA sequence. This sequence is an 813-bp fragment [nucleotide 232-1044 of SEQ ID No. 2] from clone pS12, deposited in the EMBL database under Accession number X60716, isolated from a human colon carcinoma cell line Colo320 [deposited under Accession No. ATCC CCL 220.1]. The 813-bp alphoid fragment can be obtained from the pS12 clone by nucleic acid amplification using synthetic primers, each of which contains an EcoRI site, as follows:

GGGGAATTCAT TGGGATGTTT CAGTTGA forward primer [SEQ ID No. 4]

CGAAAGTCCCC CTTAGGAGAT CTTAAGGA reverse primer [SEQ ID No. 5].

Digestion of the amplified product with EcoRI results in a fragment with EcoRI ends that includes the human α -satellite sequence. This sequence is inserted into pTEMPUD in place of the EcoRI fragment that
 5 contains the mouse major satellite to yield pHASPUD.

Vector pHASPUD was linearized with BglII and used to transform EJ30 (human fibroblast) cells by scrape loading. Twenty-seven puromycin-resistant transformant strains were obtained.

b. pTEMPHu3

10 In pTEMPHu3, the mouse major satellite sequence is replaced by the 3kb human chromosome 3-specific α -satellite from D3Z1 [deposited under ATCC Accession No. 85434; see, also Yrokov (1989) Cytogenet. Cell Genet. 51:1114].

15 5. Use of the pTEMPHU3 to induce amplification on human chromosome #3

Each human chromosome contains unique chromosome-specific aliphoid sequence. Thus, pTEMPHU3, which is targeted to the chromosome 3-specific α -satellite, can be introduced into human cells under selective conditions, whereby large-scale amplification of the
 20 chromosome 3 centromeric region and production of a *de novo* chromosome ensues. Such induced large-scale amplification provides a means for inducing *de novo* chromosome formation and also for in vivo cloning of defined human chromosome fragments up to megabase size.

For example, the break-point in human chromosome 3 is on the
 25 short arm near the centromere. This region is involved in renal cell carcinoma formation. By targeting pTEMPHu3 to this region, the induced large-scale amplification may contain this region, which can then be cloned using the bacterial and yeast markers in the pTEMPHu3 vector.

The pTEMPHu3 cloning vector allows not only selection for
 30 homologous recombinants, but also direct cloning of the integration site in YACS. This vector can also be used to target human chromosome 3,

preferably with a deleted short arm, in a mouse-human monochromosomal microcell hybrid line. Homologous recombinants can be screened by nucleic acid amplification (PCR), and amplification can be screened by DNA hybridization, Southern hybridization, and *in situ* hybridization. The amplified region can be cloned into a YAC. This vector and these methods also permit a functional analysis of cloned chromosome regions by reintroducing the cloned amplified region into mammalian cells.

B. Preparation of libraries in YAC vectors for cloning of centromeres and identification of functional chromosomal units

- Another method that may be used to obtain smaller-sized functional mammalian artificial chromosome units and to clone centromeric DNA involves screening of mammalian DNA YAC vector-based libraries and functional analysis of potential positive clones in a transgenic mouse model system. A mammalian DNA library is prepared in a YAC vector, such as YRT2 [see Schedl *et al.* (1993) Nuc. Acids Res. 21:4783-4787], which contains the murine tyrosinase gene. The library is screened for hybridization to mammalian telomere and centromere sequence probes. Positive clones are isolated and microinjected into pronuclei of fertilized oocytes of NMRI/Han mice following standard techniques. The embryos are then transferred into NMRI/Han foster mothers. Expression of the tyrosinase gene in transgenic offspring confers an identifiable phenotype (pigmentation). The clones that give rise to tyrosinase-expressing transgenic mice are thus confirmed as containing functional mammalian artificial chromosome units.
- Alternatively, fragments of SATACs may be introduced into the YAC vectors and then introduced into pronuclei of fertilized oocytes of NMRI/Han mice following standard techniques as above. The clones that give rise to tyrosinase-expressing transgenic mice are thus confirmed as containing functional mammalian artificial chromosome units, particularly centromeres.

C. Incorporation of Heterologous Genes into Mammalian Artificial Chromosomes through The Use of Homology Targeting Vectors

As described above, the use of mammalian artificial chromosomes for expression of heterologous genes obviates certain negative effects

- 5 that may result from random integration of heterologous plasmid DNA into the recipient cell genome. An essential feature of the mammalian artificial chromosome that makes it a useful tool in avoiding the negative effects of random integration is its presence as an extra-genomic gene source in recipient cells. Accordingly, methods of specific, targeted incorporation
- 10 of heterologous genes exclusively into the mammalian artificial chromosome, without extraneous random integration into the genome of recipient cells, are desired for heterologous gene expression from a mammalian artificial chromosome.

- One means of achieving site-specific integration of heterologous
- 15 genes into artificial chromosomes is through the use of homology targeting vectors. The heterologous gene of interest is subcloned into a targeting vector which contains nucleic acid sequences that are homologous to nucleotides present in the artificial chromosome. The vector is then introduced into cells containing the artificial chromosome
 - 20 for specific site-directed integration into the artificial chromosome through a recombination event at sites of homology between the vector and the chromosome. The homology targeting vectors may also contain selectable markers for ease of identifying cells that have incorporated the vector into the artificial chromosome as well as lethal selection genes that
 - 25 are expressed only upon extraneous integration of the vector into the recipient cell genome. Two exemplary homology targeting vectors, Δ CF-7 and p Δ CF-7-DTA, are described below.

1. Construction of Vector Δ CF-7

- Vector Δ CF-7 contains the cystic fibrosis transmembrane
- 30 conductance regulator [CFTR] gene as an exemplary therapeutic molecule-encoding nucleic acid that may be incorporated into mammalian artificial

chromosomes for use in gene therapy applications. This vector, which also contains the puromycin-resistance gene as a selectable marker, as well as the Saccharomyces cerevisiae ura3 gene [orotidine-5-phosphate decarboxylase], was constructed in a series of steps as follows.

5 a. Construction of pURA

Plasmid pURA was prepared by ligating a 2.6-kb Sall/XhoI fragment from the yeast artificial chromosome vector pYAC5 [Sigma; see also Burke et al. (1987) Science 236:806-812 for a description of YAC vectors as well as GenBank Accession no. U01086 for the complete
 10 sequence of pYAC5] containing the S. cerevisiae ura3 gene with a 3.3-kb Sall/SmaI fragment of pHyg [see, e.g., U.S. Patent Nos. 4,997,764, 4,686,186 and 5,162,215, and the description above]. Prior to ligation the XhoI end was treated with Klenow polymerase for blunt end ligation to the SmaI end of the 3.3 kb fragment of pHyg. Thus, pURA contains
 15 the S. cerevisiae ura3 gene, and the E. coli ColE1 origin of replication and the ampicillin-resistance gene. The uraE gene is included to provide a means to recover the integrated construct from a mammalian cell as a YAC clone.

b. Construction of pUP2

20 Plasmid pURA was digested with Sall and ligated to a 1.5-kb Sall fragment of pCEPUR. Plasmid pCEPUR is produced by ligating the 1.1 kb SnaBI-NhaI fragment of pBabe-puro [Morgenstern et al. (1990) Nucl. Acids Res. 18:3587-3596; provided by Dr. L. Székely (Microbiology and Tumorbiology Center, Karolinska Institutet, Stockholm); see, also
 25 Tonghua et al. (1995) Chin. Med. J. (Beijing, Engl. Ed.) 108:653-659; Couto et al. (1994) Infect. Immun. 62:2375-2378; Dunckley et al. (1992) FEBS Lett. 296:128-34; French et al. (1995) Anal. Biochem. 228:354-355; Liu et al. (1995) Blood 85:1095-1103; International PCT application Nos. WO 9520044; WO 9500178, and WO 9419456] to the NheI-NruI
 30 fragment of pCEP4 [Invitrogen].

The resulting plasmid, pUP2, contains all the elements of pURA plus the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal from pCEPUR.

c. Construction of pUP-CFTR

- 5 The intermediate plasmid pUP-CFTR was generated in order to combine the elements of pUP2 into a plasmid along with the CFTR gene. First, a 4.5-kb Sall fragment of pCMV-CFTR that contains the CFTR-encoding DNA [see, also, Riordan *et al.* (1989) Science 245:1066-1073, U.S. Patent No. 5,240,846, and Genbank Accession no. M28668
- 10 for the sequence of the CFTR gene] containing the CFTR gene only was ligated to XhoI-digested pCEP4 [Invitrogen and also described herein] in order to insert the CFTR gene in the multiple cloning site of the Epstein Barr virus-based (EBV) vector pCEP4 [Invitrogen, San Diego, CA; see also Yates *et al.* (1985) Nature 313:812-815; see, also U.S. Patent No.
- 15 5,468,615] between the CMV promoter and SV40 polyadenylation signal. The resulting plasmid was designated pCEP-CFTR. Plasmid pCEP-CFTR was then digested with Sall and the 5.8-kb fragment containing the CFTR gene flanked by the CMV promoter and SV40 polyadenylation signal was ligated to Sall-digested pUP2 to generate pUP-CFTR. Thus, pUP-CFTR
- 20 contains all elements of pUP2 plus the CFTR gene linked to the CMV promoter and SV40 polyadenylation signal.

d. Construction of λ CF-7

- Plasmid pUP-CFTR was then linearized by partial digestion with EcoRI and the 13 kb fragment containing the CFTR gene was ligated
- 25 with EcoRI-digested Charon 4A λ [see Blattner *et al.* (1977) Science 196:161; Williams and Blattner (1979) J. Virol. 29:555 and Sambrook *et al.* (1989) Molecular Cloning, A Laboratory Manual, Second Ed., Cold Spring Harbor Laboratory Press, Volume 1, Section 2.18, for descriptions of Charon 4A λ]. The resulting vector, λ CF8, contains the Charon 4A λ
- 30 bacteriophage left arm, the CFTR gene linked to the CMV promoter and SV40 polyadenylation signal, the ura3 gene, the puromycin-resistance

gene linked to the SV40 promoter and polyadenylation signal, the thymidine kinase promoter [TK], the ColE1 origin of replicaton, the ampicillin resistance gene and the Charon 4A λ bacteriophage right arm.

- The λ CF8 construct was then digested with XhoI and the resulting 27.1
 5 kb was ligated to the 0.4kb XhoI/EcoRI fragment of pJBP86 [described below], containing the SV40 polyA signal and the EcoRI-digested Charon 4A λ right arm. The resulting vector λ CF-7 contains the Charon 4A λ left arm, the CFTR encoding DNA linked to the CMV promoter and SV40 polyA signal, the ura3 gene, the puromycin resistance gene linked to the
 10 SV40 promoter and polyA signal and the Charon 4A λ right arm. The λ DNA fragments provide encode sequences homologous to nucleotides present in the exemplary artificial chromosomes.

- The vector is then introduced into cells containing the artificial chromosomes exemplified herein. Accordingly, when the linear λ CF-7
 15 vector is introduced into megachromosome-carrying fusion cell lines, such as described herein, it will be specifically integrated into the megachromosome through recombination between the homologous bacteriophage λ sequences of the vector and the artificial chromosome.

2. Construction of Vector λ CF-7-DTA

- 20 Vector λ CF-7-DTA also contains all the elements contained in λ CF-7, but additionally contains a lethal selection marker, the diphtheria toxin-A (DT-A) gene as well as the ampicillin-resistance gene and an origin of replication. This vector was constructed in a series of steps as follows.

a. Construction of pJBP86

- 25 Plasmid pJBP86 was used in the construction of λ CF-7, above. A 1.5-kb Sall fragment of pCEPUR containing the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal was ligated to HindIII-digested pJB8 [see, e.g., Ish-Horowitz et al. (1981) Nucleic Acids Res. 9:2989-2998; available from ATCC as Accession No. 37074;
 30 commercially available from Amersham, Arlington Heights, IL]. Prior to

ligation the Sall ends of the 1.5 kb fragment of pCEPUR and th4 HindIII linearized pJB8 ends were treated with Klenow polymerase. The resulting vector pJB86 contains the puromycin resistance gene linked to the SV40 promoter and polyA signal, the 1.8 kb COS region of Charon 4A,

- 5 the ColE1 origin of replication and the ampicillin resistance gene.

b. Construction of pMEP-DTA

- A 1.1-kb XhoI/Sall fragment of pMC1-DT-A [see, e.g., Maxwell *et al.* (1986) Cancer Res. 46:4660-4666] containing the diphtheria toxin-A gene was ligated to XhoI-digested pMEP4 [Invitrogen, San Diego, CA] to
- 10 generate pMEP-DTA. To produce pMC1-DT-A, the coding region of the DTA gene was isolated as a 800 bp PstI/HindIII fragment from p2249-1 and inserted into pMC1neopolyA [pMC1 available from Stratagene] in place of the neo gene and under the control of the TK promoter. The resulting construct pMC1DT-A was digested with HindIII, the ends filled
- 15 by Klenow and Sall linkers were ligated to produce a 1061 bp TK-DTA gene cassette with an XhoI end [5'] and a Sall end containing the 270 bp TK promoter and the ~790 bp DT-A fragment. This fragment was ligated into XhoI-digested pMEP4 .

- Plasmid pMEP-DTA thus contains the DT-A gene linked to the TK
- 20 promoter and SV40, ColE1 origin of replication and the ampicillin-resistance gene.

c. Construction of pJB83-DTA9

- Plasmid pJB8 was digested with HindIII and ClaI and ligated with an oligonucleotide [see SEQ ID NOs. 7 and 8 for the sense and
- 25 antisense strands of the oligonucleotide, respectively] to generate pJB83. The oligonucleotide that was ligated to ClaI/HindIII-digested pJB8 contained the recognition sites of SwaI, PacI and SrfI restriction endonucleases. These sites will permit ready linearization of the pICF-7-DTA construct.

- Next, a 1.4-kb XhoI/Sall fragment of pMEP-DTA, containing the DT-A gene was ligated to Sall-digested pJB83 to generate pJB83-DTA9.

d. Construction of λ CF-7-DTA

The 12-bp overhangs of λ CF-7 were removed by Mung bean nuclease and subsequent T4 polymerase treatments. The resulting 41.1-kb linear λ CF-7 vector was then ligated to pFB83-DTA9 which had been digested with ClaI and treated with T4 polymerase. The resulting vector, λ CF-7-DTA, contains all the elements of λ CF-7 as well as the DT-A gene linked to the TK promoter and the SV40 polyadenylation signal, the 1.8 kb Charon 4A λ COS region, the ampicillin-resistance gene [from pJB83-DTA9] and the Col E1 origin of replication [from pJB83-DT9A].

10 D. Targeting vectors using luciferase markers: Plasmid pMCT-RUC

Plasmid pMCT-RUC [14kbp] was constructed for site-specific targeting of the Renilla luciferase [see, e.g., U.S. Patent Nos. 5,292,658 and 5,418,155 for a description of DNA encoding *Renilla* luciferase, and plasmid pTZrLuc-1, which can provide the starting material for construction of such vectors] gene to a mammalian artificial chromosome. The relevant features of this plasmid are the *Renilla* luciferase gene under transcriptional control of the human cytomegalovirus immediate-early gene enhancer/promoter; the hygromycin-resistance gene a, positive selectable marker, under the transcriptional control of the thymidine kinase promoter. In particular, this plasmid contains plasmid pAG60 [see, e.g., U.S. Patent Nos. 5,118,620, 5,021,344, 5,063,162 and 4,946,952; see, also Colbert-Garapin et al. (1981) *J. Mol. Biol.* 150:1-14], which includes DNA (*i.e.*, the neomycin-resistance gene) homologous to the minichromosome, as well as the *Renilla* and hygromycin-resistance genes, the HSV-tk gene under control of the tk promoter as a negative selectable marker for homologous recombination, and a unique HpaI site for linearizing the plasmid.

This construct was introduced, via calcium phosphate transfection, into EC3/7C5 cells [see, Lorenz et al. (1996) *J. Biolum. Chemilum.* 11:31-37]. The EC3/7C5 cells were maintained as a monolayer [see, Gluzman (1981) *Cell* 23:175-183]. Cells at 50% confluency in 100 mm Petri

dishes were used for calcium phosphate transfection [see, Harper et al. (1981) Chromosoma 83:431-439] using 10 μ g of linearized pMCT-RUC per plate. Colonies originating from single transfected cells were isolated and maintained in F-12 medium containing hygromycin (300 μ g/mL) and

5 10% fetal bovine serum. Cells were grown in 100 mm Petri dishes prior to the *Renilla* luciferase assay.

The *Renilla* luciferase assay was performed [see, e.g., Matthews et al. (1977) Biochemistry 16:85-91]. Hygromycin-resistant cell lines obtained after transfection of EC3/7C5 cells with linearized plasmid

10 pMCT-RUC ["B" cell lines] were grown to 100% confluency for measurements of light emission in vivo and in vitro. Light emission was measured in vivo after about 30 generations as follows: growth medium was removed and replaced by 1 mL RPMI 1640 containing coelenterazine [1 mmol/L final concentration]. Light emission from cells was then

15 visualized by placing the Petri dishes in a low light video image analyzer [Hamamatsu Argus-100]. An image was formed after 5 min. of photon accumulation using 100% sensitivity of the photon counting tube. For measuring light emission in vitro, cells were trypsinized and harvested from one Petri dish, pelleted, resuspended in 1 mL assay buffer [0.5 mol/L

20 NaCl, 1 mmol/L EDTA, 0.1 mol/L potassium phosphate, pH 7.4] and sonicated on ice for 10 s. Lysates were then assayed in a Turner TD-20e luminometer for 10 s after rapid injection of 0.5 mL of 1 mmol/L coelenterazine, and the average value of light emission was recorded as LU [1 LU = 1.6×10^6 hu/s for this instrument].

25 Independent cell lines of EC3/7C5 cells transfected with linearized plasmid pMCT-RUC showed different levels of *Renilla* luciferase activity. Similar differences in light emission were observed when measurements were performed on lysates of the same cell lines. This variation in light emission was probably due to a position effect resulting from the random

30 integration of plasmid pMCT-RUC into the mouse genome, since

enrichment for site targeting of the luciferase gene was not performed in this experiment.

To obtain transfectant populations enriched in cells in which the luciferase gene had integrated into the minichromosome, transfected cells were grown in the presence of ganciclovir. This negative selection medium selects against cells in which the added pMCT-RUC plasmid integrated into the host EC3/7C5 genome. This selection thereby enriches the surviving transfectant population with cells containing pMCT-RUC in the minichromosome. The cells surviving this selection were evaluated in luciferase assays which revealed a more uniform level of luciferase expression. Additionally, the results of *in situ* hybridization assays indicated that the *Renilla* luciferase gene was contained in the minichromosome in these cells, which further indicates successful targeting of pMCT-RUC into the minichromosome.

Plasmid pNEM-1, a variant of pMCT-RUC which also contains λ DNA to provide an extended region of homology to the minichromosome [see, other targeting vectors, below], was also used to transfect EC3/7C5 cells. Site-directed targeting of the *Renilla* luciferase gene and the hygromycin-resistance gene in pNEM-1 to the minichromosome in the recipient EC3/7C5 cells was achieved. This was verified by DNA amplification analysis and by *in situ* hybridization. Additionally, luciferase gene expression was confirmed in luciferase assays of the transfectants.

E. Protein secretion targeting vectors

Isolation of heterologous proteins produced intracellularly in mammalian cell expression systems requires cell disruption under potentially harsh conditions and purification of the recombinant protein from cellular contaminants. The process of protein isolation may be greatly facilitated by secretion of the recombinantly produced protein into the extracellular medium where there are fewer contaminants to remove during purification. Therefore, secretion targeting vectors have been constructed for use with the mammalian artificial chromosome system.

A useful model vector for demonstrating production and secretion of heterologous protein in mammalian cells contains DNA encoding a readily detectable reporter protein fused to an efficient secretion signal that directs transport of the protein to the cell membrane and secretion of the protein from the cell. Vectors pLNCX-ILRUC and pLNCX-ILRUC1, described below, are examples of such vectors. These vectors contain DNA encoding an interleukin-2 (IL2) signal peptide-Renilla reniformis luciferase fusion protein. The IL-2 signal peptide [encoded by the sequence set forth in SEQ ID No. 9] directs secretion of the luciferase protein, to which it is linked, from mammalian cells. Upon secretion from the host mammalian cell, the IL-2 signal peptide is cleaved from the fusion protein to deliver mature, active, luciferase protein to the extracellular medium. Successful production and secretion of this heterologous protein can be readily detected by performing luciferase assays which measure the light emitted upon exposure of the medium to the bioluminescent luciferin substrate of the luciferase enzyme. Thus, this feature will be useful when artificial chromosomes are used for gene therapy. The presence of a functional artificial chromosome carrying an IL-Ruc fusion with the accompanying therapeutic genes will be readily monitored. Body fluids or tissues can be sampled and tested for luciferase expression by adding luciferin and appropriate cofactors and observing the bioluminescence.

1. Construction of Protein Secretion Vector pLNCX-ILRUC

Vector pLNCX-ILRUC contains a human IL-2 signal peptide-R. reniformis fusion gene linked to the human cytomegalovirus (CMV) immediate early promoter for constitutive expression of the gene in mammalian cells. The construct was prepared as follows.

a. Preparation of the IL-2 signal sequence-encoding DNA

A 69-bp DNA fragment containing DNA encoding the human IL-2 signal peptide was obtained through nucleic acid amplification, using appropriate primers for IL-2, of an HEK 293 cell line [see, e.g., U.S.

Patent No. 4,518,584 for an IL-2 encoding DNA; see, also SEQ ID No. 9; the IL-2 gene and corresponding amino acid sequence is also provided in the Genbank Sequence Database as accession nos. K02056 and J00264]. The signal peptide includes the first 20 amino acids shown in

5 the translations provided in both of these Genbank entries and in SEQ ID NO. 9. The corresponding nucleotide sequence encoding the first 20 amino acids is also provided in these entries [see, e.g., nucleotides 293-52 of accession no. K02056 and nucleotides 478-537 of accession no. J00264], as well as in SEQ ID NO. 9. The amplification primers included

10 an EcoRI site [GAATTC] for subcloning of the DNA fragment after ligation into pGEMT [Promega]. The forward primer is set forth in SEQ ID No. 11 and the sequence of the reverse primer is set forth in SEQ ID No. 12.

TTTGAATTCATGTACAGGATGCAACTCCTG forward [SEQ ID No. 11]
TTTGAATTCAGTAGGTGCACTGTTTGTGAC reverse [SEQ ID No. 12]

15 **b. Preparation of the R. reniformis luciferase-encoding DNA**

The initial source of the R. reniformis luciferase gene was plasmid pLXSN-RUC. Vector pLXSN [see, e.g., U.S. Patent Nos. 5,324,655, 5,470,730, 5,468,634, 5,358,866 and Miller et al. (1989)

20 Biotechniques 7:980] is a retroviral vector capable of expressing heterologous DNA under the transcriptional control of the retroviral LTR; it also contains the neomycin-resistance gene operatively linked for expression to the SV40 early region promoter. The R. reniformis luciferase gene was obtained from plasmid pTZrLuc-1 [see, e.g., U.S.

25 Patent No. 5,292,658; see also the Genbank Sequence Database accession no. M63501; and see also Lorenz et al. (1991) Proc. Natl. Acad. Sci. U.S.A. 88:4438-4442] and is shown as SEQ ID NO. 10. The 0.97 kb EcoRI/SmaI fragment of pTZrLuc-1 contains the coding region of the *Renilla* luciferase-encoding DNA. Vector pLXSN was digested with

30 and ligated with the luciferase gene contained on a pLXSN-RUC, which contains the luciferase gene located operably linked to the viral LTR and

upstream of the SV40 promoter, which directs expression of the neomycin-resistance gene.

c. **Fusion of DNA encoding the IL-2 Signal Peptide and the R. reniformis Luciferase Gene to Yield pLXSN-ILRUC**

5

The pGEMT vector containing the IL-2 signal peptide-encoding DNA described in 1.a. above was digested with EcoRI, and the resulting fragment encoding the signal peptide was ligated to EcoRI-digested pLXSN-RUC. The resulting plasmid, called pLXSN-ILRUC, contains the IL-

10 2 signal peptide-encoding DNA located immediately upstream of the R. reniformis gene in pLXSN-RUC. Plasmid pLXSN-ILRUC was then used as a template for nucleic acid amplification of the fusion gene in order to add a SmaI site at the 3' end of the fusion gene. The amplification product was subcloned into linearized [EcoRI/SmaI-digested] pGEMT [Promega] to
15 generate ILRUC-pGEMT.

d. **Introduction of the Fusion Gene into a Vector Containing Control Elements for Expression in Mammalian Cells**

Plasmid ILRUC-pGEMT was digested with KspI and SmaI to
20 release a fragment containing the IL-2 signal peptide-luciferase fusion gene which was ligated to HpaI-digested pLNCX. Vector pLNCX [see, e.g., U.S. Patent Nos. 5,324,655 and 5,457,182; see, also Miller and Rosman (1989) Biotechniques 7:980-990] is a retroviral vector for expressing heterologous DNA under the control of the CMV promoter; it
25 also contains the neomycin-resistance gene under the transcriptional control of a viral promoter. The vector resulting from the ligation reaction was designated pLNCX-ILRUC. Vector pLNCX-ILRUC contains the IL-2 signal peptide-luciferase fusion gene located immediately downstream of the CMV promoter and upstream of the viral 3' LTR and polyadenylation
30 signal in pLNCX. This arrangement provides for expression of the fusion gene under the control of the CMV promoter. Placement of the heterologous protein-encoding DNA [i.e., the luciferase gene] in operative

linkage with the IL-2 signal peptide-encoding DNA provides for expression of the fusion in mammalian cells transfected with the vector such that the heterologous protein is secreted from the host cell into the extracellular medium.

5 **2. Construction of Protein Secretion Targeting Vector pLNCX-ILRUC1**

Vector pLNCX-ILRUC may be modified so that it can be used to introduce the IL-2 signal peptide-luciferase fusion gene into a mammalian artificial chromosome in a host cell. To facilitate specific incorporation of
10 the pLNCX-ILRUC expression vector into a mammalian artificial chromosome, nucleic acid sequences that are homologous to nucleotides present in the artificial chromosome are added to the vector to permit site directed recombination.

Exemplary artificial chromosomes described herein contain λ phage
15 DNA. Therefore, protein secretion targeting vector pLNCX-ILRUC1 was prepared by addition of λ phage DNA [from Charon 4A arms] to produce the secretion vector pLNCX-ILRUC.

3. **Expression and Secretion of R. reniformis Luciferase from Mammalian Cells**

20 **a. Expression of R. reniformis Luciferase Using pLNCX-ILRUC**

Mammalian cells [LMTK⁺ from the ATCC] were transiently transfected with vector pLNCX-ILRUC [$\sim 10 \mu\text{g}$] by electroporation [BIORAD, performed according to the manufacturer's instructions]. Stable
25 transfectants produced by growth in G418 for neo selection have also been prepared.

Transfectants were grown and then analyzed for expression of luciferase. To determine whether active luciferase was secreted from the transfected cells, culture media were assayed for luciferase by addition of
30 coelentrastazine [see, e.g., Matthews *et al.* (1977) *Biochemistry* 16:85-91].

The results of these assays establish that vector pLNCX-ILRUC is capable of providing constitutive expression of heterologous DNA in

- mammalian host cells. Furthermore, the results demonstrate that the human IL-2 signal peptide is capable of directing secretion of proteins fused to the C-terminus of the peptide. Additionally, these data demonstrate that the R. reniformis luciferase protein is a highly effective
- 5 reporter molecule, which is stable in a mammalian cell environment, and forms the basis of a sensitive, facile assay for gene expression.

b. ***Renilla reniformis* luciferase appears to be secreted from LMTK⁻ cells.**

(i) ***Renilla* luciferase assay of cell pellets**

- 10 The following cells were tested:
- cells with no vector: LMTK⁻ cells without vector as a negative control;
 - cells transfected with pLNCX only;
 - cells transfected with RUC-pLNCX [*Renilla* luciferase gene in pLNCX vector];
- 15 cells transfected with pLNCX-ILRUC [vector containing the IL-2 leader sequence + *Renilla* luciferase fusion gene in pLNCX vector].
- Forty-eight hours after electroporation, the cells and culture medium were collected. The cell pellet from 4 plates of cells was
- 20 resuspended in 1 ml assay buffer and was lysed by sonication. Two hundred μ l of the resuspended cell pellet was used for each assay for luciferase activity [see, e.g., Matthews et al. (1977) Biochemistry 16:85-91]. The assay was repeated three times and the average bioluminescence measurement was obtained.
- 25 The results showed that there was relatively low background bioluminescence in the cells transformed with pLNCX or the negative control cells; there was a low level observed in the cell pellet from cells containing the vector with the IL-2 leader sequence-luciferase gene fusion and more than 5000 RLU in the sample from cells containing RUC-pLNCX.

(ii) **Renilla luciferase assay of cell medium**

Forty milliliters of medium from 4 plates of cells were harvested and spun down. Two hundred microliters of medium was used for each luciferase activity assay. The assay was repeated several times and the
 5 average bioluminescence measurement was obtained. These results showed that a relatively high level of bioluminescence was detected in the cell medium from cells transformed with pLNCX-ILRUC; about 10-fold lower levels [slightly above the background levels in medium from cells with no vector or transfected with pLNCX only] was detected in the cells
 10 transfected with RUC-pLNCX.

(iii) **conclusions**

The results of these experiments demonstrated that Renilla luciferase appears to be secreted from LMTK⁻ cells under the direction of the IL-2 signal peptide. The medium from cells transfected with Renilla
 15 luciferase-encoding DNA linked to the DNA encoding the IL-2 secretion signal had substantially higher levels of Renilla luciferase activity than controls or cells containing luciferase-encoding DNA without the signal peptide-encoding DNA. Also, the differences between the controls and cells containing luciferase encoding-DNA demonstrate that the luciferase
 20 activity is specifically from luciferase, not from a non-specific reaction. In addition, the results from the medium of RUC-pLNCX transfected cells, which is similar to background, show that the luciferase activity in the medium does not come from cell lysis, but from secreted luciferase.

c. **Expression of R. reniformis Luciferase Using pLNCX-ILRUC λ**

To express the IL-2 signal peptide-R. reniformis fusion gene from an mammalian artificial chromosome, vector pLNCX-ILRUC λ is targeted for site-specific integration into a mammalian artificial chromosome through homologous recombination of the λ DNA sequences contained in the
 30 chromosome and the vector. This is accomplished by introduction of pLNCX-ILRUC λ into either a fusion cell line harboring mammalian artificial

chromosomes or mammalian host cells that contain mammalian artificial chromosomes. If the vector is introduced into a fusion cell line harboring the artificial chromosomes, for example through microinjection of the vector or transfection of the fusion cell line with the vector, the cells are

- 5 then grown under selective conditions. The artificial chromosomes, which have incorporated vector pLNCX-ILRUC λ , are isolated from the surviving cells, using purification procedures as described above, and then injected into the mammalian host cells.

- 10 Alternatively, the mammalian host cells may first be injected with mammalian artificial chromosomes which have been isolated from a fusion cell line. The host cells are then transfected with vector pLNCX-ILRUC λ and grown.

The recombinant host cells are then assayed for luciferase expression as described above.

15 F. Other targeting vectors

These vectors, which are based on vector pMCT-RUC, rely on positive and negative selection to insure insertion and selection for the double recombinants. A single crossover results in incorporation of the DT-A, which kills the cell, double crossover recombinations delete the DT-
20 1 gene.

1. Plasmid pNEM1 contains:

- | | |
|-------|--|
| DT-A: | Diphtheria toxin gene (negative selectable marker) |
| Hyg: | Hygromycin gene (positive selectable marker) |
| ruc: | <u>Renilla</u> luciferase gene (non-selectable marker) |
| 25 1: | LTR-MMTV promoter |
| 2: | TK promoter |
| 3: | CMV promoter |
| MMR: | Homology region (plasmid pAG60) |

2. plasmid pNEM-2 and -3 are similar to pNEM 1 except for

- 30 different negative selectable markers:
pNEM-1: diphtheria toxin gene as "-" selectable marker

- pNEM-2: hygromycin antisense gene as "—" selectable marker
- pNEM-3: thymidine kinase HSV-1 gene as "—" selectable marker
3. Plasmid - λ DNA based homology:
- pNEM-1: base vector
- 5 pNEM-2: base vector containing p5 = gene
- 1: LTR MMTV promoter
- 2: SV40 promoter
- 3: CMV promoter
- 4: μ TIIA promoter (metallothionein gene promoter)
- 10 — homology region (plasmid pAG60)
- λ L.A. and λ R.A. homology regions for λ left and right arms (λ gt-WES).

EXAMPLE 13

Microinjection of mammalian cells with plasmid DNA

- 15 These procedures will be used to microinject MACs into eukaryotic cells, including mammalian and insect cells.

- The microinjection technique is based on the use of small glass capillaries as a delivery system into cells and has been used for introduction of DNA fragments into nuclei [see, e.g., Chalfie et al. (1994) Science 263:802-804]. It allows the transfer of almost any type of molecules, e.g., hormones, proteins, DNA and RNA, into either the cytoplasm or nuclei of recipient cells. This technique has no cell type restriction and is more efficient than other methods, including Ca^{2+} —mediated gene transfer and liposome-mediated gene transfer.
- 20
- 25 About 20-30% of the injected cells become successfully transformed.

- Microinjection is performed under a phase-contrast microscope. A glass microcapillary, prefilled with the DNA sample, is directed into a cell to be injected with the aid of a micromanipulator. An appropriate sample volume (1-10 pl) is transferred into the cell by gentle air pressure exerted
- 30 by a transjector connected to the capillary. Recipient cells are grown on

glass slides imprinted with numbered squares for convenient localization of the injected cells.

a. Materials and equipment

- Nunclon tissue culture dishes 35 x 10 mm, mouse cell line EC3/7C5
- 5 Plasmid DNA pCH110 [Pharmacia], Purified Green Florescent Protein (GFP) [GFPs from *Aequorea* and *Renilla* have been purified and also DNA encoding GFPs has been cloned; see, e.g., Prasher et al. (1992) Gene 111:229-233; International PCT Application No. WO 95/07463, which is based on U.S. application Serial No. 08/119,678 and U.S. application
- 10 Serial No. 08/192,274], ZEISS Axiovert 100 microscope, Eppendorf transjector 5246, Eppendorf micromanipulator 5171, Eppendorf Cellocate coverslips, Eppendorf microloaders, Eppendorf femtotips and other standard equipment.

b. Protocol for injecting

- 15 (1) Fibroblast cells are grown in 35 mm tissue culture dishes (37° C, 5% CO₂) until the cell density reaches 80% confluency. The dishes are removed from the incubator and medium is added to about a 5 mm depth.
- (2) The dish is placed onto the dish holder
- 20 and the cells observed with 10 x objective; the focus is desirably above the cell surface.
- (3) Plasmid or chromosomal DNA solution [1 ng/ μ l] and GFP protein solution are further purified by centrifuging the DNA sample at a force sufficient to remove any particular debris [typically
- 25 about 10,000 rpm for 10 minutes in a microcentrifuge].
- (4) Two 2 μ l of the DNA solution (1 ng/ μ l) is loaded into a microcapillary with an Eppendorf microloader. During loading, the loader is inserted to the tip end of the microcapillary. GFP (1 mg/ml) is loaded with the same procedure.

(5) The protecting sheath is removed from the microcapillary and the microcapillary is fixed onto the capillary holder connected with the micromanipulator.

(6) The capillary tip is lowered to the surface of the medium and is focussed on the cells gradually until the tip of the capillary reaches the surface of a cell. The capillary is lowered further so that the it is inserted into the cell. Various parameters, such as the level of the capillary, the time and pressure, are determined for the particular equipment. For example, using the fibroblast cell line C5 and the above-noted equipment, the best conditions are: injection time 0.4 second, pressure 80 psi. DNA can then be automatically injected into the nuclei of the cells.

(7) After injection, the cells are returned to the incubator, and incubated for about 18-24 hours.

(8) After incubation the number of transformants can be determined by a suitable method, which depends upon the selection marker. For example, if green fluorescent protein is used, the assay can be performed using UV light source and fluorescent filter set at 0-24 hours after injection. If β -gal-containing DNA, such as DNA-derived from pHC110, has been injected, then the transformants can be assayed for β -gal.

(c) Detection of β -galactosidase in cells injected with plasmid DNA

The medium is removed from the culture plate and the cells are fixed by addition of 5 ml of fixation Solution I: (1% glutaraldehyde; 0.1 M sodium phosphate buffer, pH 7.0; 1 mM $MgCl_2$), and incubated for 15 minutes at 37° C. Fixation Solution I is replaced with 5 ml of X-gal Solution II: [0.2% X-gal, 10 mM sodium phosphate buffer (pH 7.0), 150 mM NaCl, 1 mM $MgCl_2$, 3.3 mM $K_4Fe(CN)_6 \cdot H_2O$, 3.3 mM $K_3Fe(CN)_6$], and the plates are incubated for 30-60 minutes at 37° C. The X-gal solution is

removed and 2 ml of 70% glycerol is added to each dish. Blue stained cells are identified under a light microscope.

This method will be used to introduce a MAC, particularly the MAC with the anti-HIV megachromosome, to produce a mouse model for anti-

5 HIV activity.

EXAMPLE 14

Transgenic (non-human) animals

Transgenic (non-human) animals can be generated that express heterologous genes which confer desired traits, e.g., disease resistance,

10 in the animals. A transgenic mouse is prepared to serve as a model of a disease-resistant animal. Genes that encode vaccines or that encode therapeutic molecules can be introduced into embryos or ES cells to produce animals that express the gene product and thereby are resistant to or less susceptible to a particular disorder.

15 The mammalian artificial megachromosome and others of the artificial chromosomes, particularly the SATACs, can be used to generate transgenic (non-human) animals, including mammals and birds, that stably express genes conferring desired traits, such as genes conferring resistance to pathogenic viruses. The artificial chromosomes can also be
20 used to produce transgenic (non-human) animals, such as pigs, that can produce immunologically humanized organs for xenotransplantation.

For example, transgenic mice containing a transgene encoding an anti-HIV ribozyme provide a useful model for the development of stable transgenic (non-human) animals using these methods. The artificial
25 chromosomes can be used to produce transgenic (non-human) animals, particularly, cows, goats, mice, oxen, camels, pigs and sheep, that produce the proteins of interest in their milk; and to produce transgenic chickens and other egg-producing fowl, that produce therapeutic proteins or other proteins of interest in their eggs. For example, use of mammary

gland-specific promoters for expression of heterologous DNA in milk is known [see, e.g. U.S. Patent No. 4,873,316]. In particular, a milk-specific promoter or a promoter, preferably linked to a milk-specific signal peptide, specifically activated in mammary tissue is operatively

5 linked to the DNA of interest, thereby providing expression of that DNA sequence in milk.

1. **Development of Control Transgenic Mice Expressing Anti-HIV Ribozyme**

Control transgenic mice are generated in order to compare stability

10 and amounts of transgene expression in mice developed using transgene DNA carried on a vector (control mice) with expression in mice developed using transgenes carried in an artificial megachromosome.

a. **Development of Control Transgenic Mice Expressing β -galactosidase**

One set of control transgenic mice was generated by microinjection

of mouse embryos with the β -galactosidase gene alone. The microinjection procedure used to introduce the plasmid DNA into the mouse embryos is as described in Example 13, but modified for use with embryos [see, e.g., Hogan et al. (1994) *Manipulating the Mouse Embryo*,

20 *A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, see, especially pages 255-264 and Appendix 3]. Fertilized mouse embryos [Strain CB6 obtained from Charles River Co.] were injected with 1 ng of plasmid pCH110 (Pharmacia) which had been linearized by digestion with BamHI. This plasmid contains the β -

25 galactosidase gene linked to the SV40 late promoter. The β -galactosidase gene product provides a readily detectable marker for successful transgene expression. Furthermore, these control mice provide confirmation of the microinjection procedure used to introduce the plasmid into the embryos. Additionally, because the megachromosome

30 that is transferred to the mouse embryos in the model system (see below) also contains the β -galactosidase gene, the control transgenic mice that have been generated by injection of pCH110 into embryos serve as an

analogous system for comparison of heterologous gene expression from a plasmid versus from a gene carried on an artificial chromosome.

After injection, the embryos are cultured in modified HTF medium under 5% CO₂ at 37°C for one day until they divide to form two cells.

- 5 The two-cell embryos are then implanted into surrogate mother female mice [for procedures see, Manipulating the Mouse Embryo, A Laboratory Manual (1994) Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 127 et seq.].

- b. **Development of Control Transgenic Mice Expressing Anti-HIV Ribozyme**

- 10 One set of anti-HIV ribozyme gene-containing control transgenic mice was generated by microinjection of mouse embryos with plasmid pCEPUR-132 which contains three different genes: (1) DNA encoding an anti-HIV ribozyme, (2) the puromycin-resistance gene and (3) the
- 15 hygromycin-resistance gene. Plasmid pCEPUR-132 was constructed by ligating portions of plasmid pCEP-132 containing the anti-HIV ribozyme gene (referred to as ribozyme D by Chang et al. [(1990) Clin. Biotech. 2:23-31]; see also U.S. Patent No. 5,144,019 to Rossi et al., particularly Figure 4 of the patent) and the hygromycin-resistance gene with a portion
- 20 of plasmid pCEPUR containing the puromycin-resistance gene.

- 25 Plasmid pCEP-132 was constructed as follows. Vector pCEP4 (Invitrogen, San Diego, CA; see also Yates et al. (1985) Nature 313:812-815) was digested with XhoI which cleaves in the multiple cloning site region of the vector. This ~10.4-kb vector contains the hygromycin-resistance gene linked to the thymidine kinase gene promoter and
- polyadenylation signal, as well as the ampicillin-resistance gene and ColE1 origin of replication and EBNA-1 (Epstein-Barr virus nuclear antigen) genes and OriP. The multiple cloning site is flanked by the cytomegalovirus promoter and SV40 polyadenylation signal.

- 30 XhoI-digested pCEP4 was ligated with a fragment obtained by digestion of plasmid 132 (see Example 4 for a description of this plasmid)

with XhoI and Sall. This XhoI/Sall fragment contains the anti-HIV ribozyme gene linked at the 3' end to the SV40 polyadenylation signal. The plasmid resulting from this ligation was designated pCEP-132. Thus, in effect, pCEP-132 comprises pCEP4 with the anti-HIV ribozyme gene

5 and SV40 polyadenylation signal inserted in the multiple cloning site for CMV promoter-driven expression of the anti-HIV ribozyme gene.

To generate pCEPUR-132, pCEP-132 was ligated with a fragment of pCEPUR. pCEPUR was prepared by ligating a 7.7-kb fragment generated upon NheI/NruI digestion of pCEP4 with a 1.1-kb NheI/SnaBI

10 fragment of pBabe [see Morgenstern and Land (1990) Nucleic Acids Res. 18:3587-3596 for a description of pBabe] that contains the puromycin-resistance gene linked at the 5' end to the SV40 promoter. Thus, pCEPUR is made up of the ampicillin-resistance and EBNA1 genes, as well as the ColE1 and OriP elements from pCEP4 and the puromycin-resistance

15 gene from pBabe. The puromycin-resistance gene in pCEPUR is flanked by the SV40 promoter (from pBabe) at the 5' end and the SV40 polyadenylation signal (from pCEP4) at the 3' end.

Plasmid pCEPUR was digested with XhoI and Sall and the fragment containing the puromycin-resistance gene linked at the 5' end to the

20 SV40 promoter was ligated with XhoI-digested pCEP-132 to yield the ~12.1-kb plasmid designated pCEPUR-132. Thus, pCEPUR-132, in effect, comprises pCEP-132 with puromycin-resistance gene and SV40 promoter inserted at the XhoI site. The main elements of pCEPUR-132 are the hygromycin-resistance gene linked to the thymidine kinase

25 promoter and polyadenylation signal, the anti-HIV ribozyme gene linked to the CMV promoter and SV40 polyadenylation signal, and the puromycin-resistance gene linked to the SV40 promoter and polyadenylation signal. The plasmid also contains the ampicillin-resistance and EBNA1 genes and the ColE1 origin of replication and OriP.

30 Zygotes were prepared from (C57BL/6JxCBA/J) F1 female mice [see, e.g., Manipulating the Mouse Embryo, A Laboratory Manual (1994)

- Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 429], which had been previously mated with a (C57BL/6JxCBA/J) F1 male. The male pronuclei of these F2 zygotes were injected [see, Manipulating the Mouse Embryo, A Laboratory Manual 5 (1994) Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY] with pCEPUR-132 (~3 µg/ml), which had been linearized by digestion with NruI. The injected eggs were then implanted in surrogate mother female mice for development into transgenic offspring.
- 10 These primary carrier offspring were analyzed (as described below) for the presence of the transgene in DNA isolated from tail cells. Seven carrier mice that contained transgenes in their tail cells (but that may not carry the transgene in all their cells, i.e., they may be chimeric) were allowed to mate to produce non-chimeric or germ-line heterozygotes. The 15 heterozygotes were, in turn, crossed to generate homozygote transgenic offspring.

2. Development of Model Transgenic Mice Using Mammalian Artificial Chromosomes

- Fertilized mouse embryos are microinjected (as described above) 20 with megachromosomes (1-10 pL containing 0-1 chromosomes/pL) isolated from fusion cell line G3D5 or H1D3 (described above). The megachromosomes are isolated as described herein. Megachromosomes isolated from either cell line carry the anti-HIV ribozyme (ribozyme D) gene as well as the hygromycin-resistance and β -galactosidase genes.
- 25 The injected embryos are then developed into transgenic mice as described above.

- Alternatively, the megachromosome-containing cell line G3D5* or H1D3* is fused with mouse embryonic stem cells [see, e.g., U.S. Patent No. 5,453,357, commercially available; see Manipulating the Mouse 30 Embryo, A Laboratory Manual (1994) Hogan et al., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pages 253-289]

following standard procedures see also, e.g., Guide to Techniques in Mouse Development in Methods in Enzymology Vol. 25, Wassarman and De Pamphilis, eds. (1993), pages 803-932]. (It is also possible to deliver isolated megachromosomes into embryonic stem cells using the Microcell

5 procedure [such as that described above].) The stem cells are cultured in the presence of a fibroblast [e.g., STO fibroblasts that are resistant to hygromycin and puromycin]. Cells of the resultant fusion cell line, which contains megachromosomes carrying the transgenes [i.e., anti-HIV ribozyme, hygromycin-resistance and β -galactosidase genes], are then

10 transplanted into mouse blastocysts, which are in turn implanted into a surrogate mother female mouse where development into a transgenic mouse will occur.

Mice generated by this method are chimeric; the transgenes will be expressed in only certain areas of the mouse, e.g., the head, and thus

15 may not be expressed in all cells.

3. Analysis of Transgenic Mice for Transgene Expression

Beginning when the transgenic mice, generated as described above, are three-to-four weeks old, they can be analyzed for stable expression of the transgenes that were transferred into the embryos [or

20 fertilized eggs] from which they develop. The transgenic mice may be analyzed in several ways as follows.

a. Analysis of Cells Obtained from the Transgenic Mice

Cell samples [e.g., spleen, liver and kidney cells, lymphocytes, tail

25 cells] are obtained from the transgenic mice. Any cells may be tested for transgene expression. If, however, the mice are chimeras generated by microinjection of fertilized eggs or by fusion of embryonic stem cells with megachromosome-containing cells, only cells from areas of the mouse that carry the transgene are expected to express the transgene. If the

30 cells survive growth on hygromycin [or hygromycin and puromycin or neomycin, if the cells are obtained from mice generated by transfer of

both antibiotic-resistance genes], this is one indication that they are stably expressing the transgenes. RNA isolated from the cells according to standard methods may also be analyzed by northern blot procedures to determine if the cells express transcripts that hybridize to nucleic acid probes based on the antibiotic-resistance genes. Additionally, cells obtained from the transgenic mice may also be analyzed for β -galactosidase expression using standard assays for this marker enzyme [for example, by direct staining of the product of a reaction involving β -galactosidase and the X-gal substrate, see, e.g., Jones (1986) EMBO 5:3133-3142, or by measurement of β -galactosidase activity, see, e.g., Miller (1972) in Experiments in Molecular Genetics pp. 352-355, Cold Spring Harbor Press]. Analysis of β -galactosidase expression is particularly used to evaluate transgene expression in cells obtained from control transgenic mice in which the only transgene transferred into the embryo was the β -galactosidase gene.

Stable expression of the anti-HIV ribozyme gene in cells obtained from the transgenic mice may be evaluated in several ways. First, DNA isolated from the cells according to standard procedures may be subjected to nucleic acid amplification using primers corresponding to the ribozyme gene sequence. If the gene is contained within the cells, an amplified product of pre-determined size is detected upon hybridization of the reaction mixture to a nucleic acid probe based on the ribozyme gene sequence. Furthermore, DNA isolated from the cells may be analyzed using Southern blot methods for hybridization to such a nucleic acid probe. Second, RNA isolated from the cells may be subjected to northern blot hybridization to determine if the cells express RNA that hybridizes to nucleic acid probes based on the ribozyme gene. Third, the cells may be analyzed for the presence of anti-HIV ribozyme activity as described, for example, in Chang et al. (1990) Clin. Biotech. 2:23-31. In this analysis, RNA isolated from the cells is mixed with radioactively labeled HIV gag target RNA which can be obtained by in vitro transcription of gag gene

template under reaction conditions favorable to in vitro cleavage of the gag target, such as those described in Chang et al. (1990) Clin. Biotech. 2:23-31. After the reaction has been stopped, the mixture is analyzed by gel electrophoresis to determine if cleavage products smaller in size than
 5 the whole template are detected; presence of such cleavage fragments is indicative of the presence of stably expressed ribozyme.

b. Analysis of Whole Transgenic Mice

Whole transgenic mice that have been generated by transfer of the anti-HIV ribozyme gene [as well as selection and marker genes] into
 10 embryos or fertilized eggs can additionally be analyzed for transgene expression by challenging the mice with infection with HIV. It is possible for mice to be infected with HIV upon intraperitoneal injection with high-producing HIV-infected U937 cells [see, e.g., Locardi et al. (1992) J. Virol. 66:1649-1654]. Successful infection may be confirmed by analysis
 15 of DNA isolated from cells, such as peripheral blood mononuclear cells, obtained from transgenic mice that have been injected with HIV-infected human cells. The DNA of infected transgenic mice cells will contain HIV-specific gag and env sequences, as demonstrated by, for example, nucleic acid amplification using HIV-specific primers. If the cells also
 20 stably express the anti-HIV ribozyme, then analysis of RNA extracts of the cells should reveal the smaller gag fragments arising by cleavage of the gag transcript by the ribozyme.

Additionally, the transgenic mice carrying the anti-HIV ribozyme gene can be crossed with transgenic mice expressing human CD4 (i.e.,
 25 the cellular receptor for HIV) [see Gillespie et al. (1993) Mol. Cell. Biol. 13:2952-2958; Hanna et al. (1994) Mol. Cell. Biol. 14:1084-1094; and Yeung et al. (1994) J. Exp. Med. 180:1911-1920, for a description of transgenic mice expressing human CD4]. The offspring of these crossed transgenic mice expressing both the CD4 and anti-HIV ribozyme
 30 transgenes should be more resistant to infection [as a result of a

reduction in the levels of active HIV in the cells] than mice expressing CD4 alone [without expressing anti-HIV ribozyme].

4. Development of transgenic chickens using artificial chromosomes

- 5 The development of transgenic chickens has many applications in the improvement of domestic poultry, an agricultural species of commercial significance, such as disease resistance genes and genes encoding therapeutic proteins. It appears that efforts in the area of chicken transgenesis have been hampered due to difficulty in achieving
- 10 stable expression of transgenes in chicken cells using conventional methods of gene transfer via random introduction into recipient cells. Artificial chromosomes are, therefore, particularly useful in the development of transgenic chickens because they provide for stable maintenance of transgenes in host cells.
- 15 **a. Preparation of artificial chromosomes for introduction of transgenes into recipient chicken cells**
- (i) Mammalian artificial chromosomes**
- Mammalian artificial chromosomes, such as the SATACs and minichromosomes described herein, can be modified to incorporate
- 20 detectable reporter genes and/or transgenes of interest for use in developing transgenic chickens. Alternatively, chicken-specific artificial chromosomes can be constructed using the methods herein. In particular, chicken artificial chromosomes [CACs] can be prepared using the methods herein for preparing MACs; or, as described above, the chicken
- 25 libraries can be introduced into MACs provided herein and the resulting MACs introduced into chicken cells and those that are functional in chicken cells selected.

- As described in Examples 4 and 7, and elsewhere herein, artificial chromosome-containing mouse LMTK-derived cell lines, or
- 30 minichromosome-containing cell lines, as well as hybrids thereof, can be transfected with selected DNA to generate MACs [or CACs] that have

integrated the foreign DNA for functional expression of heterologous genes contained within the DNA.

- To generate MACs or CACs containing transgenes to be expressed in chicken cells, the MAC-containing cell lines may be transfected with
- 5 DNA that includes λ DNA and transgenes of interest operably linked to a promoter that is capable of driving expression of genes in chicken cells. Alternatively, the minichromosomes or MACs [or CACs], produced as described above, can be isolated and introduced into cells, followed by targeted integration of selected DNA. Vectors for targeted integration are
- 10 provided herein or can be constructed as described herein.

- Promoters of interest include constitutive, inducible and tissue (or cell)-specific promoters known to those of skill in the art to promote expression of genes in chicken cells. For example, expression of the lacZ gene in chicken blastodermal cells and primary chicken fibroblasts has
- 15 been demonstrated using a mouse heat-shock protein 68 (*hsp 68*) promoter [phspPTlacZpA; see Brazolot *et al.* (1991) Mol. Reprod. Devel. 30:304-312], a Zn²⁺-inducible chicken metallothionein (cMt) promoter [pCBcMtlacZ; see Brazolot *et al.* (1991) Mol. Reprod. Devel. 30:304-312], the constitutive Rous sarcoma virus and chicken β -actin promoters in
- 20 tandem [pmivZ; see Brazolot *et al.* (1991) Mol. Reprod. Devel. 30:304-312] and the constitutive cytomegalovirus (CMV) promoter. Of particular interest herein are egg-specific promoters that are derived from genes, such as ovalbumin and lysozyme, that are expressed in eggs.

- The choice of promoter will depend on a variety of factors,
- 25 including, for example, whether the transgene product is to be expressed throughout the transgenic chicken or restricted to certain locations, such as the egg. Cell-specific promoters functional in chickens include the steroid-responsive promoter of the egg ovalbumin protein-encoding gene [see Gaub *et al.* (1987) EMBO J. 6:2313-2320; Tora *et al.* (1988) EMBO J. 7:3771-3778; Park *et al.* (1995) Biochem. Mol. Biol. Int. (Australia) 36:811-816].
- 30

(ii) Chicken artificial chromosomes

Additionally, chicken artificial chromosomes may be generated using methods described herein. For example, chicken cells, such as primary chicken fibroblasts [see Brazolot *et al.* (1991) Mol. Reprod. Devel. 30:304-312], may be transfected with DNA that encodes a selectable marker [such as a protein that confers resistance to antibiotics] and that includes DNA (such as chicken satellite DNA) that targets the introduced DNA to the pericentric region of the endogenous chicken chromosomes.

- Transfectants that survive growth on selection medium are then analyzed, using methods described herein, for the presence of artificial chromosomes, including minichromosomes, and particularly SATACs. An artificial chromosome-containing transfectant cell line may then be transfected with DNA encoding the transgene of interest [fused to an appropriate promoter] along with DNA that targets the foreign DNA to the chicken artificial chromosome.

b. Introduction of artificial chromosomes carrying transgenes of interest into recipient chicken cells

- Cell lines containing artificial chromosomes that harbor transgene(s) of interest (i.e., donor cells) may be fused with recipient chicken cells in order to transfer the chromosomes into the recipient cells. Alternatively, the artificial chromosomes may be isolated from the donor cells, for example, using methods described herein [see, *e.g.*, Example 10], and directly introduced into recipient cells.

- Exemplary chicken recipient cell lines include, but are not limited to, stage X blastoderm cells [see, *e.g.*, Brazolot *et al.* (1991) Mol. Reprod. Dev. 30:304-312; Etches *et al.* (1993) Poultry Sci. 72:882-889; Petitte *et al.* (1990) Development 108:185-189] and chick zygotes [see, *e.g.*, Love *et al.* (1994) Biotechnology 12:60-63].

- For example, microcell fusion is one method for introduction of artificial chromosomes into avian cells [see, *e.g.*, Dieken *et al.* (1996) Nature Genet. 12:174-182 for methods of fusing microcells with DT40

chicken pre-B cells]. In this method, microcells are prepared [for example, using procedures described in Example 1.A.5] from the artificial chromosome-containing cell lines and fused with chicken recipient cells.

- Isolated artificial chromosomes may be directly introduced into
- 5 chicken recipient cell lines through, for example, lipid-mediated carrier systems, such as lipofection procedures [see, e.g., Brazolot et al. (1991) Mol. Reprod. Dev. 30:304-312] or direct microinjection. Microinjection is generally preferred for introduction of the artificial chromosomes into chicken zygotes [see, e.g., Love et al. (1994) Biotechnology 12:60-63].

10 **c. Development of transgenic chickens**

- Transgenic chickens may be developed by injecting recipient Stage X blastoderm cells (which have received the artificial chromosomes) into embryos at a similar stage of development [see, e.g., Etches et al. (1993) Poultry Sci. 72:882-889; Petitte et al. (1990) Development 108:185-189;
- 15 and Carsience et al. (1993) Development 117: 669-675]. The recipient chicken embryos within the shell are candled and allowed to hatch to yield a germline chimeric chicken that will express the transgene(s) in some of its cells.

- Alternatively, the artificial chromosomes may be introduced into
- 20 chick zygotes, for example through direct microinjection [see, e.g., Love et al. (1994) Biotechnology 12:60-63], which thereby are incorporated into at least a portion of the cells in the chicken. Inclusion of a tissue-specific promoter, such as an egg-specific promoter, will ensure appropriate expression of operatively-linked heterologous DNA.

- 25 The DNA of interest may also be introduced into a minichromosome, by methods provided herein. The minichromosome may either be one provided herein, or one generated in chicken cells using the methods herein. The heterologous DNA will be introduced using a targeting vector, such as those provided herein, or constructed as
- 30 provided herein.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

007241022 4420456
00000000 00000000

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Hadlaczky, Gyula
Szalay, Aladar
- (ii) TITLE OF THE INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
- (iii) NUMBER OF SEQUENCES: 34
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Heller Ehrman White & McAuliffe
(B) STREET: 4250 Executive Square, 7th Floor
(C) CITY: La Jolla
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92037
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 28-NOV-2000
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/835,682
(B) FILING DATE: 10-APR-1997
(C) CLASSIFICATION:
- (viii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/695,191
(B) FILING DATE: 07-AUG-1996
(C) CLASSIFICATION:
- (ix) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/682,080
(B) FILING DATE: 15-JUL-1996
(C) CLASSIFICATION:
- (x) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/629,822
(B) FILING DATE: 10-APR-1996
(C) CLASSIFICATION:
- (xi) ATTORNEY/AGENT INFORMATION:
(A) NAME: Seidman, Stephanie L
(B) REGISTRATION NUMBER: 33,779
(C) REFERENCE/DOCKET NUMBER: 24601-402F
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 858-450-8403
(B) TELEFAX: 858-587-5360
(C) TELEX:
- (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1293 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:
 (ix) FEATURE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCATCA	TTTTTCANGT	CCTCAAGTGG	ATGTTTCTCA	TTTNCATGA	TTTTAAGTTT	60
TCTCGCCATA	TTCCTGTGTC	TACAGTGTGC	ATTTCCTCAT	TTTNCACGTT	TTNCAGTGAT	120
TTCGTCAATT	TCAAGTCCTC	AAGTGGATGT	TTCTCATTIN	CCATGAATTT	CAGTTTTCCTN	180
GCCATATTTC	ACGTCTCTACA	GNGGACATTT	CTAAATTTNC	CACCTTTTTTC	AGTTTTCCTC	240
GCCATATTTC	ACGTCTCTAAA	ATGTGTAATT	CTCGTTTNC	GTGATTTTCA	GTTTTTCCTC	300
CAGATTCCAG	GTCTATAAAT	GTGCATTCTC	CATTNNCAC	GTTTTTCAGT	GTTCCTGTC	360
TTTTTTCAG	TCCGCAAGTG	GATGTTTCTC	ATTNNCCATG	ATTNNCAGTT	TTCCTGNAAT	420
ATTCCATGTC	CTACAATGAT	CATTTTAAAT	TTTCCACCTT	TTCAATTTTC	CACGCCATAT	480
TTCAATGCTC	AAAGTGATATA	TTTCTCCTTT	TCCGCGAATT	CTAGTTTCTC	CGCCATATTC	540
CAGGTCTCAT	AGGTGTGCAAT	CCTCATTTT	CACCTTTTTC	ACTGATTTTC	TCATTTTTC	600
AGTCGTCAAC	TGATCTTTTC	TAATTTTCCA	TGATTTTCAG	TTATCTGTCT	ATATTCCATG	660
TCCTACAGTG	AGATTTTCTA	AAATTTTCCA	CTTTTTCAT	TTTTCTCGAC	ATATTTCAGC	720
TGCTAAAGTG	TGATTTTCTC	ATTTCCTG	ATTTCAGTT	TTCTCGGCAT	ATTCCAGGTC	780
CTAATAGTGT	GCATTTTCTA	TTTTTCAGT	TTTTTCAGTA	TTTCTGCTAT	TTTTCCAGTT	840
GTCAAGGGGA	TGTTTCTCAT	TTTTCCATG	TGTCAGTTT	CTTGCTATAT	TCCATGCTCT	900
ACAGTGGCAT	TCTTAATAT	TATACCTTT	TCAGTTTTC	TCACCATAT	TCACGCTCTA	960
AAGTATATAT	TCTTAATAT	CCTGATTTT	CAGTTTCTCT	GCATATTCC	AGGTCTCTACA	1020
GTGTGCATTT	CTCATTTTTC	ACGTTTTC	GTAATTTCT	CATTTTTPAA	GGCCTCAAT	1080
GGATTTTCT	CATTTTTCAT	GATTTTCAGT	TTTTCTGAGG	TATACCATGT	CCTACAGTGG	1140
ACATTTCTAA	ATTATCCACC	TTTTCAGTT	TTTCTCGGCG	ACATTTTCAG	TCCTAAAGTG	1200
TGTAATTCAT	ATTTCAGT	ATTTCAGT	TTCTCGCCAT	ATTCCAGGAC	CTACAGTGTG	1260
CATTTCCTCAT	TTTTTCAGTT	TTTTTCAGT	TTC			1293

- (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1044 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:
 (ix) FEATURE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AGGCCTATGG	TGAAAAAGGA	AATATCTTCC	CCTGAAACT	AGACAGAAGG	ATTCTCAGAA	60
TCTTATTGTT	GATGTGCGCC	CCTCAACTAA	CAGTGTGAA	GCITTTCTTT	GATAGAGCAG	120
TTTTGAAACA	CTCTTTTGT	AAAATCTGCA	AGAGGATATT	TGGATAGCTT	TGAGGATTC	180
CGTTGGAAC	GGGATTGTCT	TCATATAAAC	CCTAGACAGA	AGCATTTCTA	GAAGCTTCAT	240
TGGGATGTTT	CAGTTGAAGT	CACAGTGTG	AACAGTCCC	TTTTCATAGAG	CAGGTTTGAA	300
ACACTCTTTT	TTGTAGTATC	TGGAAGTGG	CATTGTGAGC	GATCTCAGA	CTCGGTGAA	360
AAAGGAANA	TCTTCCAATA	AAAGCTAGAT	AGAGGCAATG	TCGAGAACCT	TTTTCATGAT	420
GTATCTACTC	AGCTAACAGA	GTGTAACCTT	CCTTGTAGAG	AGCACTTTTG	AAACACTCTT	480
TTTTGTGAAT	CTCAAGTGG	ATAATGCTCT	AGCTTTGAGC	ATTTCTGTGG	GAACAGGGAT	540
TACATATAAA	AAGCAGACAG	CGCATCTCC	AGAAACTTCT	TTGTGATGTT	TGATTCAGG	600
TCACAGAGTT	GAACATCTCC	TTTCATAGAG	CAGGTTTGAA	ACACACTTTT	TGATGTATCT	660

GGATGTGGAC	ATTTCAGCGC	CTTTCAGGCC	TAAGGTGAAA	AGGAAATATC	TTCCCTTGAA	720
AACTAGACAG	AAGCATTCTC	AGAACTTAT	TTGTGATGTG	CGCCCTCAAC	TAAACAGTGT	780
GAAGCTTTCT	TTTGTATAGAG	GCAAGTTTGA	AACACTCTTT	TGTGGAATCT	GCAAGTGGAT	840
ATTTGTCTAG	CTTTAGGAGT	TTCTTTGGAA	ACGGGATATC	ATATAAAAAA	CAGACAGCAG	900
CATTCCCGAG	ATCTTGTGTT	TGATGTTTGC	ATTCAAGTCA	CAGAGTTGAA	CATTCCCTTT	960
CAGAGACGAG	GTTTGAACAC	TCTTTTATA	GTATCTGGAT	TGGGACATTT	GGAGCGCTTT	1020
CAGGGGGGAT	CCTCTAGAAT	TCCT				1044

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2492 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTGCAGCTGG	GGGTCTCCAA	TCAGGCAGGG	GGCCCTTACT	ACTCAGATGG	GGTGGCCGAG	60
TAGGGGAAGG	GGGTGCGAGC	TGCATGAGTG	GACACAGCTG	TAGGACTACC	TGGGGGCTGT	120
GGATCTATGG	GGGTGCGGAG	AAGCCCACTG	ACAGTGCCTA	GAAGAGACAA	GGTGGCCTGA	180
GAGGGTCTGA	GCAACATAGA	GCTGCCCATG	TTGGGGCCAG	GTCTCAAGCA	GGAAGTGAAG	240
AATGGGACAG	GCTCTCAGAT	ACTCTACTCA	GTAGCCAGGA	TAGCAAGGAG	GGCTTGGGGT	300
TGCTATCTCT	GGGTTCAACC	CCCCAGTTG	AAGGCCCTGG	GGGAGATGGT	CCCAGGACAT	360
ATTACAATGG	ACACAGGAGG	TGGGGACACC	TGGAGTCACC	AAACAAACCC	ATGCCAAGAG	420
AGACCATATG	TAGGGGTGTC	CAGTCCAGCC	CTCTGACTGA	GCTGCATTGT	TCAAATCCAA	480
AGGGCCCCCTG	CTGCCACCTA	GTGGCTGATG	GCATCCACAT	GACCCTGGGC	CACACGCGTT	540
TAGGTCCTCT	TGGAAGACCA	AGATCCTTGT	TACATTGAAC	GACTCCTAAA	GAGCAGAGA	600
TTTCCACCTA	TTCCGAACAA	TCACATAAAA	TCCATCCTGG	AAAAAGCCTG	GGGGATGGCA	660
CTAAGGCTAG	GGATAGGGTG	GGATGAAGAT	TATAGTTACA	GTAAGGGGTT	TAGGGTTAGG	720
GATCAACGTT	GGTTAGGAGT	TAGGGATACA	GTAAGGGTACC	GGTAGGGTTA	GGGGTTAGGG	780
TTAGGGGGTTA	GGGTTAGGGT	TAGGGTTAGG	GTTAGGGTTA	GGGGTTAGGG	GTTAGGGTTA	840
GGGTTAGGTT	TTGGGGTGGC	GTATTTTGGT	CTTATACGCT	GTGTTCCACT	GGCAATGAAA	900
AGAGTCTCTG	TTTTCCTCTC	AGCAATTGTG	CATTTTAAAA	AGAGTTTAGC	AATTCTAACA	960
GATATAGAAC	AGCTGTGCTA	TCTCAATTGT	GTTTTCAATT	GTAACACAT	TGTGGTTTCA	1020
ATGTTGTTAC	TTGCCATCTG	TAGATCTTCT	TTGCGTGAGG	GTCTGTGTCA	GATGTGTGTG	1080
CATTTCCTTG	NTTTNGGCTG	TTTAACTTAT	TGTTTAGTTT	TAATAATTTT	TTTATATATT	1140
GAAGACAAT	CTTTCTCAGA	TGTGTAATTT	CAAAATTTT	TTCAATATTT	GGCTTGCCTT	1200
TGTCCTCAAC	AAGGTCCTCT	CAGAGATAAC	TTAAATATAA	GAATATCCCA	CTGTCACTTC	1260
TTTTGTGTAT	ATCTACCTTT	TGTGTCAATT	GTTAAATATC	ATTACCAAAC	CCAAAGGCGAC	1320
ATAGCTTTTC	TTCTATTGTT	TCTTCTAGAA	ATTTGTATAT	TTTTGCAATT	TTAGTGTAAG	1380
GATGATTTTG	AGTGATTGTT	TGTGTAAGTT	TGTAAGTTT	CGTCTATATC	CATATCAITT	1440
CTTATGGTTT	CCAATTAACT	TTTCCCTCAC	TATTTTGGGG	AAAGACACAG	GATAGTGGGC	1500
TTTGTTAGAG	TAGATAGGTA	GCTAGACATG	AACAGGAGGG	GGCCTCTGG	AAAGGGGAAA	1560
CTGTGGGAGG	CTGCACCTGG	AGGACCACCA	AAAATTCCAA	TATTAGTAGC	ATCTCTAGTG	1620
GCGGATGAGA	TGGGCACCTG	TCAATTGTGG	TGAGGAGGGA	AAAGAGGTCC	TATCTAGAAA	1680
GAAACTCCCT	AGAACTCTCT	TGAAGATGNC	CCAATCATTC	ACTCTGCAAT	AAATATGTCA	1740
GAATATTGCT	AGCTACATGT	TGATAAGGNN	AAAGGGGAGA	TCCTTAAGTG	AAACCTTGCA	1800
CCATTAAGTAC	AGATTATAGG	AGAGAAGGAC	ATTCAAAAGA	CGGGAAGGAG	GAGTGTATGC	1860
ACGTGATGCG	TGTCAAGTGT	CCTGGGATGG	CGGGAAGGAG	GCTGGTCCCA	CAGTGAATTC	1920
GCTATTGATCA	CCACACATAT	ACCTCAACCA	ACAGTGAAGA	GGTCCACAAA	GCTTAAGTGG	1980
GGCAGATTGG	GGAGCTAAGG	CAGTAGCAGG	AAACCCAGAC	AAAGAAACAA	GGTGGAGACT	2040
TGAGACAGAG	CAGGAAATGT	GAGAGAAATC	AAATATAAAT	CTCCCTGACA	GGACTCTTAG	2100
GCTGTTTAAAT	GCATCGCTCA	GTTCACCTCC	TCCCTATTTT	TCTACAATTA	ACTCTTTTCA	2160
CTGTGTTTCT	TTTCAATGAA	GTTCCTCTGG	ATCTTTGTAT	TGCTCTCTGG	TGAAATGTTT	2220
CTTCCCAAGT	TAAACAAGAA	CTGGGACATC	AGCTCTCCCC	AGTAATAGCT	CCGTTTCACT	2280

TTGAATTAC	AGAACTGATG	GGCTTAATAA	CTGGCGCTCT	GACTTTAGTG	GTGCAGGAGG	2340
CCGTCAACACC	GGGACCAAGA	GTGCCCTGCC	TAGTCCCCAT	CTGCCCGCAG	GTGGCGGCTG	2400
CCTCGACACT	GACAGCAATA	GGGTCCGGCA	GTGTCCCCAG	CTGCCACGAG	GGGGCGTAGC	2460
ACGACTACAC	TGTGAGCAAG	AGGGCCCTGC	AG			2492

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGGGAATTCA TTGGGATGTT TCAGTTGA

28

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGAAAGTCCC CCCTAGGAGA TCTTAAGGA

29

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCGCTTAATA CTCTGATGAG TCCGTGAGGA CGAAACGCTC TCGCACC

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:
 (ix) FEATURE:
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGATTTAAT TAATTAAGCC CGGGC

25

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:
 (ix) FEATURE:
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TAAATTAAAT TAATTCGGGC CCGTCGA

27

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 69 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
 (D) OTHER INFORMATION IL-2 signal sequence
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATG TAC AGG ATG CAA CTC CTG TCT TGC ATT GCA CTA AGT CTT GCA CTT
 Met Tyr Arg Met Gln Leu Leu Ser Cys Ile Ala Leu Ser Leu Ala Leu

GTC ACA AAC AGT GCA CCT ACT
 Val Thr Asn Ser Ala Pro Thr

48

69

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 945 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence

(B) LOCATION: 1...942

(D) OTHER INFORMATION: Renilla Reinformis Luciferase

(x) PUBLICATION INFORMATION:

PATENT NO.: 5,418,155

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGC TTA AAG ATG ACT TCG AAA GTT TAT GAT CCA GAA CAA AGG AAA CGG	48
Ser Leu Lys Met Thr Ser Lys Val Tyr Asp Pro Glu Gln Arg Lys Arg	
1 5 10 15	
ATG ATA ACT GGT CCG CAG TGG TGG GCC AGA TGT AAA CAA ATG AAT GTT	96
Met Ile Thr Gly Pro Gln Trp Trp Ala Arg Cys Lys Gln Met Asn Val	
20 25 30	
CTT GAT TCA TTT ATT AAT TAT TAT GAT TCA GAA AAA CAT GCA GAA AAT	144
Leu Asp Ser Phe Ile Asn Tyr Tyr Asp Ser Glu Lys His Ala Glu Asn	
35 40 45	
GCT GTT ATT TTT TTA CAT GGT AAC GCG GCC TCT TCT TAT TTA TGG CGA	192
Ala Val Ile Phe Leu His Gly Asn Ala Ala Ser Ser Tyr Leu Trp Arg	
50 55 60	
CAT GTT GTG CCA CAT ATT GAG CCA GTA GCG CGG TGT ATT ATA CCA GAT	240
His Val Val Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp	
65 70 75 80	
CTT ATT GGT ATG GGC AAA TCA GGC AAA TCT GGT AAT GGT TCT TAT AGG	288
Leu Ile Gly Met Gly Lys Ser Gly Lys Ser Gly Asn Gly Ser Tyr Arg	
85 90 95	
TTA CTT GAT CAT TAC AAA TAT CTT ACT GCA TGG TTG AAC TTC TTA ATT	336
Leu Leu Asp His Tyr Lys Tyr Leu Thr Ala Trp Leu Asn Phe Leu Ile	
100 105 110	
TAC CAA AGA AGA TCA TTT TTT GTC GGC CAT GAT TGG GGT GCT TGT TTG	384
Tyr Gln Arg Arg Ser Phe Phe Val Gly His Asp Trp Gly Ala Cys Leu	
115 120 125	
GCA TTT CAT TAT AGC TAT GAG CAT CAA GAT AAG ATC AAA GCA ATA GTT	432
Ala Phe His Tyr Ser Tyr Glu His Gln Asp Lys Ile Lys Ala Ile Val	
130 135 140	
CAC GCT GAA AGT GTA GTA GAT GTG ATT GAA TCA TGG GAT GAA TGG CCT	480
His Ala Glu Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro	
145 150 155 160	
GAT ATT GAA GAA GAT ATT GCG TTG ATC AAA TCT GAA GAA GGA GAA AAA	528
Asp Ile Glu Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Glu Lys	
165 170 175	
ATG GTT TTG GAG AAT AAC TTC TTC GTG GAA ACC ATG TTG CCA TCA AAA	576
Met Val Leu Glu Asn Asn Phe Phe Val Glu Thr Met Leu Pro Ser Lys	
180 185 190	
ATC ATG AGA AAG TTA GAA CCA GAA GAA TTT GCA GCA TAT CTT GAA CCA	624

Ile Met Arg	Lys Leu Glu Pro Glu Glu Phe Ala Ala Tyr Leu Glu Pro	
195	200	205
TTC AAA GAG AAA GGT GAA GTT CGT CGT CCA ACA TTA TCA TGG CCT CGT	672	
Phe Lys Glu Lys Gly Glu Val Arg Arg Pro Thr		
210	215	220
GAA ATC CCG TTA GTA AAA GGT GGT AAA CCT GAC GTT GTA CAA ATT GTT	720	
Glu Ile Pro Leu Val Lys Gly Gly Lys Pro Asp Val Val Gln Ile Val		
225	230	235
AGG AAT TAT AAT GCT TAT CTA CGT GCA AGT GAT GAT TTA CCA AAA ATG	768	
Arg Asn Tyr Asn Ala Tyr Leu Arg Ala Ser Asp Asp Leu Pro Lys Met		
245	250	255
TTT ATT GAA TCG GAT CCA GGA TTC TTT TCC AAT GCT ATT GTT GAA GGC	816	
Phe Ile Glu Ser Asp Pro Gly Phe Phe Ser Asn Ala Ile Val Glu Gly		
260	265	270
GCC AAG AAG TTT CCT AAT ACT GAA TTT GTC AAA GTA AAA GGT CTT CAT	864	
Ala Lys Lys Lys Phe Pro Asn Thr Glu Phe Val Lys Val Lys Gly Leu His		
275	280	285
TTT TCG CAA GAA GAT GCA CCT GAT GAA ATG GGA AAA TAT ATC AAA TCG	912	
Phe Ser Gln Glu Asp Ala Pro Asp Glu Met Gly Lys Tyr Ile Lys Ser		
290	295	300
TTC GTT GAG CGA GTT CTC AAA AAT GAA CAA TAA	945	
Phe Val Glu Arg Val Leu Lys Asn Glu Gln		
305	310	

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTTGAATTC A TGTACAGGAT GCAACTCCTG

30

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TTTGAATTCA GTAGGTGCAC TGTTTGTCCAC

30

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1434 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCTCCACGCA	CGTTGTGATA	TGTAGATGAT	AATCATTATC	AGAGCAGCGT	TGGGGGATAA	60
TGTCGACATT	TCCACTCCCA	ATGACGGTGA	TGTATAATGC	TCAAGTATTC	TCCTGCTTTT	120
TTACCACTAA	CTAGGAACCT	GGTTTGGCCT	TAATTCAGAC	AGCCTTGGCT	CTGTCTGGAC	180
AGGTCACGAC	GACTGACACC	ATTAACACTT	TGTCAGCCCTC	AGTGACTACA	GTCTAGATAGT	240
AACAGGCCTC	AGCTAATGTC	AGATACAGA	GAGGTCTCAT	GCTGGTTAAT	CAACTCATAG	300
ATCTTGTCCA	GATCAACACT	GATGTATTAT	GACAAATAAC	TCAGCAGGGA	TGTGAACAAA	360
AGTTTCCGGG	ATTGTGTGTT	ATTTCATTTC	AGTATGTTAA	ATTACTAGG	ACAGCTAATT	420
TGTCAAAAAG	TCTTTTTCAG	TATATGTTAC	AGAATTGGAT	GGCTGAATTT	GAACAGATCC	480
TTCCGGGAAT	GAGACTTCAG	GTCAACTCCA	CGCGCTTGGA	CTGTGCTGCT	ACCAAGGAT	540
TACCCAAATT	GATCTCTCTCA	GCATTTTCTT	TCTTTAAAAA	ATGGGTGGGA	TTAATATTAT	600
TTGGAGATAC	ACTTTTGTGT	GGATTAGTGT	TGCTTCTTTG	ATTGGTCTGT	AAGCTTAAGG	660
CCCAAACTAG	GAGAGACAAG	GTGGTTATTG	CCGAGGCGCT	TGCAGGACTA	GAACATGGAG	720
CTTCCCCIGA	TATATGCTTA	TCTATGCTTA	GGCAATAGGT	CGCTGGCCAC	TCAGCTCTTTA	780
TATCCCCACA	GGCTAGTCTC	ATTGTACGGG	ATAGAGTGAG	TGTGCTTCAG	CAGCCCGAGA	840
GAGTTGCAAG	GCTAAGCACT	GCAATGGAAA	GGCTCTGCGG	CATATATGTG	CCTATTCTAG	900
GGGGACATGT	CATCTTTTCAT	GAAGGTTTCAG	TGTCCTTAGT	CCCTTCCCCC	AGGCAAAACG	960
ACACGGGAGC	AGGTCAGGGT	TGCTCTGGGT	AAAAGCCTGT	GAGCCTGGGA	GCTAATCCTG	1020
TACATGGCTC	CTTTACCTAC	ACACTGGGGA	TTTGACCTCT	ATCTCCACTC	TCATTAAATAT	1080
GGGTGGCCCTA	TTTGCTCTTA	TTAAAAGGAA	AGGGGGAGAT	GTTGGGAGCC	GCGCCACAT	1140
TCGCCGTTAC	AAGATGGCGC	TGACAGCTGT	GTTCTAAGTG	GTAACAAAT	AATCTGCGCA	1200
TGTGCCGAGG	TGGTTCTCTC	ACTCCATGTG	CTCTGCTTC	CCCGTGACGT	CAACTCGGCC	1260
GATGGGCTGC	AGCCAACTAG	GGAGTGACAC	GTCTTAGGCG	AAGGAGAATT	CTCTTAATA	1320
GGGACGGGGT	TTCTTCTCTC	CTCTCTCTCT	TGCTTCTCTC	TCTTGCTTTT	TCGCTCTCTT	1380
GCTTCCCGTA	AAGTGATAAT	GATTATCATC	TACATATCAC	AACGTGCGTG	GAGG	1434

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1400 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CCTCCACGCA	CGTTGTGATA	TGTAGATGAT	AATCATTATC	AGAGCAGCGT	TGGGGGATAA	60
TGTCGACATT	TCCACTCCCA	ATGACGGTGA	TGTATAATGC	TCAAGTATTC	TCCTGCTTTT	120
TTACCACTAA	CTAGGAACCT	GGTTTGGCCT	TAATTCAGAC	AGCCTTGGCT	CTGTCTGGAC	180
AGGTCACGAT	ACAACATAGT	GTATTATGAC	AAATAACTCA	GCAGGGATGT	GAACAAAAGT	240

TTCCGGGATT	GCCTGTTATT	TCCATCCAGT	ATGTTAAATT	TACTAGGGCA	GCTAATTGT	300
CAAAAAGTCT	TTTCCAGTAT	ATGTTACAGA	ATTGGATGGC	TGAATTGAA	CAGATCCTTC	360
GGGAATTGAG	ACTTCCAGTTC	AACTCCACGC	GCTTGGACCT	GTCCCTGACC	AAAGGATTAC	420
CCAATTGGAT	CTCCTCAGCA	TTTTCTTTCT	TTAAAAAATG	GGTGGGATTA	ATATTATTTC	480
GAGATACACT	TTGCTGTGGA	TTAGTGTGTC	TTCTTTGATT	GCTCTGTAG	CTTAAGGCCCT	540
AAACTAGGAG	AGACAAGGTG	GTTATTGCC	AGGCCTGTGC	AGGACTAGAA	CATGGAGCTT	600
CCCTCATGAT	ATCTATGCTT	AGGCAATAGG	TCGCTGGCCA	CTCAGCTCTT	ATATCCCATG	660
AGGCTAGTCT	CATGTCACGG	GATAGAGTGA	GTGTGCTTCA	GCAGCCCGAG	AGAGTTGCAC	720
GGCTAAGCAC	TGCAATTGGA	AGGCTCTGCG	GCATATTATGA	GCCTATTCTA	GGGAGACATG	780
TCATCTTTCA	AGAAAGTTGA	GTGTCCAAGT	TCCTTCTCTC	CAGGCCAAAC	GACACGGGAG	840
CAGGTCAAGG	TTGCTCTGGG	TAAAGCCCTG	TGAGCCTAAG	AGCTAATCCT	GTACATGGCT	900
CCTTTACCTA	CACACTGGGG	ATTTGACCTC	TATCTCCACT	CTCATTAAAT	TGGGTGGCCT	960
ATTTGCTCTT	ATTTAAAGGA	AAGGGGGAGA	TGTTGGGAGC	CGCGCCCAAC	TTCCGCGCTA	1020
CAAGATGGCG	CTGACAGCTG	TGTTCTAAGT	GGTAACACAA	TAATCTGCGC	ATGCGCCGAG	1080
GGTGGTTCTT	CACCTCCATG	GCTCTGCCCT	CCCCGTGACG	TCAACTCGCG	CGATGCGCTG	1140
CAGTCAATCA	GGGAGTGACA	CGTCTTAGGC	GAAGGAAAT	TCTCTTAAAT	AGGGACGGGG	1200
TTTCGTTTTT	TCTCTCTCTT	GCTTCGCTCT	CTCTTGCTTC	TGCTCTCTCT	TTCCCTGAAG	1260
TGTAAGAATA	AAGCTTTGCC	GCAGAAGATT	CTGGTCTGTG	GTGTCTCTTC	TGCGCGGTGC	1320
TGAGAACCGC	TCTAATAACA	ATTGGTGCCG	AAACCCGGGT	GATAATGATT	ATCATCTACA	1380
TATCACCAAG	TGCCTGGAGG					1400

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1369 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CCTCCAGGCA	CGTTGTGATA	TGTAGATGAT	AATCATTATC	ACTTTACGGG	TCCTTTCACT	60
ACAACATGCC	CGAGGCCCCG	TGCTCTGGTA	ATAGATCTTT	GCTGAAAAGG	CACACACATG	120
ACACATTACT	CAAGGTGGGC	TCATCTGAGC	TGCAGATTCA	GCTTAATATG	AACTCTGGCA	180
ATTGTGTGAA	ATCATAAATC	TTCAAGTGGA	CACCTATTGC	CGACACAGAG	TGCCACCTT	240
TGGCATAATA	AACAAACACA	AATTATCTAT	TATATAAAGG	GTGTTAGAG	ATGCTTTAGA	300
ATACAAATAA	ATCATGGTAG	ATAACAGTAA	GTTGAGAGCT	TAAATTAAAT	AAAGTGATAT	360
ACCTAATAAA	AATTAATAAT	AGAAGGTGTG	AATATACTAC	AGTAGGTAAA	TTATTTCAAT	420
AATTTTATTT	CTTTCTTAAT	CCTTTATAAT	GTTTCTGCT	ATTGTCAATT	GCACATCCAT	480
ATGTTCAATT	CTTCACTGTA	ATGAAGAAAT	GTAGTAAATA	TACTTTCCGA	ACAAGTTTGA	540
TCAAATATGT	TACACTTGAT	TCCGTGTGTT	ACCTATCATT	TTATTATTAT	ATTGATTGCA	600
TTCTCTTGTT	ACTTGATATT	ATTACAAGGT	ACATATTCTT	TCCTTCAGAT	CTTCATTATA	660
CTCTAACCAT	TTTATAACAT	CTCTTTATTA	TTCAATTTCIT	TGCTGACAAA	TGAGGCACAA	720
ATGCCAGAGA	GAACCTGAGC	AGATAAGAGG	ACAAATTGCA	AGAGTCAGTT	ACCTCTCTGT	780
GTTCCTTGGA	AACTCAGTGC	CAAAATTCAG	TTGTGAGGCT	TGGCAGCATG	CACCTTTTAC	840
CAGTGCCTCC	ATCTTGTCTG	CCCTGAACAT	CAAGCTTTGC	AGACAGACAG	GCTACACTAA	900
GTGACCTGGT	CATTCAACAG	ATGCATGGTG	ATTTATTGTT	ACTTTCTATT	CCATGCTTTT	960
ACTATTCTTA	CTAGGTGCTA	GCTAGTACTG	TATTTTCGAGA	TAGAAGTTAC	TGAAGAGAAA	1020
TTACATGTTT	CTTTGATACT	CCTTGATACT	ATATAGAGTT	TTAATCAGST	TTAATCAGST	1080
CCTAGACCTT	TTCTTCACTC	TTATTAATAA	CTAAGTACAA	ATTAAGTTTA	TCCAAACAGT	1140
TACGAGTGTG	GATTTTGTGC	AGTTCCTACT	TGATAATGAT	CTAGCTTCAT	AAATCTGACA	1200
CACCTTATGG	GAATGTTTTT	GTAAATAAAA	GATTCAAGGTG	TTACTCTAGG	TCAAGAGAAAT	1260
ATTAAACATC	AGTCCCAAAAT	TACAAACCTC	AATAAAAGAT	TGTACTCTCC	AGTGGTGGA	1320
ATATAAAGTG	ATAATGATTA	TCACTCTACT	ATCAACACGT	GCGTGGAGG		1369

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22118 base pairs

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAATTCGCCCT	ATCCCTAATC	CAGATTGGTG	GAATAACTTG	GTATAGATGT	TTGTGCATTA	60
AAAACCCCTGT	AGGATTCGA	CTCTAGGTCA	CTGTTGACGA	CTGGAAACCTG	AATTGTGGCC	120
CTGACTGATA	GGTCTTGCA	CATATGCAGT	TCTGACACAGA	CAGACACAGCA	GACACACAGA	180
CAGACACAGCA	GACAGACGTT	ACAAACAAC	ACGTTGAGCC	GTGTGCCAAC	ACACACACAA	240
ACACCACCTCT	GGCCATAATT	ATTGAGGACG	TTGATTATT	ATTCTGTGTT	TGTGAGTCTG	300
CTCTGTCTGTC	TGCTGTCTGT	TCTGTCTGTC	TATCAAACCA	AAAGAAACCA	AAACAATTATG	360
CCTGCTGGCC	TGCTGCTGTC	CTACACACAGA	GAATGATT	CTTCAATCAA	TCTAAAACGA	420
CCTCTCAAGT	TTGCTTTTCT	TCTCTTTTCT	TATCTTTTCT	TTTTTTCTTT	TCTTCTTCTCT	480
TCTCTTCTCT	TCTCTTCTCT	TCTCTTCTCT	CTTCTTCTCT	TCTTCTTCTT	CTTACTTTCT	540
TTCTTTCTCT	CTTACTTTTA	TTCTTTTCTAT	ACATAGTTTC	TTAGTGTAA	CATCCCTGAC	600
TGTCCTGAAG	ACACTTTGTA	GGCCTCAATC	CTGTAAGAGC	CTTCCCTCTG	TTTTCAAAAT	660
CTGGCATGAA	TGTTGTACCT	CACATGACCC	AGCTTAGTCT	TCAAGTCTGA	GTTCATGGAA	720
AGGAGTTTCA	AGAAGACTGG	TTATATTTTT	CATTATTAT	TGCATTTTAA	TTAAATTTTA	780
ATTTCAACAA	AAGAATTATG	ACTGACCAAT	TCAGAGTCTG	CCGTTTAAAA	GCATAAGGAA	840
AAAGTAGTCA	AAAAACGTGA	GGCTGTCTGT	GGATGTCGA	GGCTGCTTTA	GGGAGCCTCG	900
TCACCATCTCT	GCACTTGCAA	ACCGGGCCAC	TAGAACCCTG	TGAAGGGAGA	AAACAAAGCG	960
ACCTGGAAAC	ATTAGGTCTAC	ATGAAGGCCA	GCCACCTCCA	CTTGTGTTGT	CGGGAGTTCA	1020
GTTAGCAGAC	AAGATGCTGC	CCATGACACAT	GTGTCTTTTC	AGCTTGGTGA	GGTCAAAGTA	1080
CAACCGAGTC	ACAGAACCAAG	GAAGTATACA	CAGTGAGTCT	CAGGTGACCC	AGAGTTTACA	1140
CAGAGAAACC	ACATCTTGAA	AAAAACAAAA	AAATAAATTA	AATAAATATA	ATTTAAAAAT	1200
ACCTGGAAAC	ATTAGGTCTAC	ATGAAGGCCA	CTTAAATCCC	AGCTCTCTTC	AGGCAGAGAT	1260
TTAAAAATAG	CCGGGAGTGA	TGGCGCATGT	GTCTGCAAA	TGAGTTCCAG	GACAGTCCAG	1320
GGGAGGATTT	CTGAGTTTGA	GGCCAGCCTG	GTCTGCAAA	AACTAAACT	AAACTAAAAA	1380
GCTATACAGA	GAACCCCTGT	CTTGAANAAT	AACTAAACT	AACTAAACT	AACTAAAAA	1440
AATATAAAT	AAAAATTTTA	AAGAATTTTA	AAAAACTACA	GAATCAAC	ATAGCCAC	1500
GAGATGGCAA	GTAACTGCAA	TCATAGCAGA	AATATTATAC	ACACACACAC	ACACAGACTC	1560
TGTCATAAAA	TCCAATGTGC	CTTCAATGATG	ATCAAAATTC	GATAGTCACT	AATACTAGAA	1620
GAATCATATG	TCTGAAAATA	AAAGCCAGAA	CCTTTTCTCT	TTTGTTTTTC	TTTTGCCCCA	1680
AGATAGGGTG	TCTCTCAGTG	TATCCCTGTC	ATCCCTGCTC	CTGCCCTGCT	GCCTGCCCTG	1740
GGTAGCCTCA	AACCTCAGAA	GTCTCTCTCT	CTTCTCTCTG	CAACCAACAC	ACCGAGTCCA	1800
CTGCGCTGCC	GCCTGCCCTCA	CTTCTCTCTC	CTTCTCTCTC	TTCTTTCTCT	TTCTTTCTCT	1860
TTTATTTCTT	TCTTTTCTCT	TTCTTTTCTC	TTCTTTTCTC	TTCTTTTCTC	TTCTTTTCTC	1920
CTTCTTTCTT	TCTTTTCTCT	TTCTTTTCTC	TTCTTTTCTC	TTCTTTTCTC	TTCTTTTCTC	1980
TGCTTATAGG	CTGCTTTGCC	CTGCTTTGCC	CTGCTTTGCC	CTGCTTTGCC	CTGCTTTGCC	2040
TCTGTAGAAAT	AAGTGAARAA	AGTGAARAA	AGTGAARAA	AGTGAARAA	AGTGAARAA	2100
ATATGCGCAG	CTGTGCAGAG	CTGTGCAGAG	CTGTGCAGAG	CTGTGCAGAG	CTGTGCAGAG	2160
TGCTTTTAT	GCACACACAG	GCACACACAG	GCACACACAG	GCACACACAG	GCACACACAG	2220
AAAGCAGGCT	GGCCTTGAC	GGCCTTGAC	GGCCTTGAC	GGCCTTGAC	GGCCTTGAC	2280
AAAGCAGGCT	GGCCTTGAC	GGCCTTGAC	GGCCTTGAC	GGCCTTGAC	GGCCTTGAC	2340
CTTTCTTTCT	TTTTCTCTCT	TTTTCTCTCT	TTTTCTCTCT	TTTTCTCTCT	TTTTCTCTCT	2400
TTTTCTTTT	TTTTTTT	TTTTTTT	TTTTTTT	TTTTTTT	TTTTTTT	2460
AATTTGCCCTA	GTCTGTCTCT	AATTTCTCTT	AAAAAAAAC	AAACAAAAA	AAACAAAAA	2520
CAGTATGTAT	GTATGTATAT	TTAGAAGAA	TACTAATCCA	TAAATACTA	TTTTTTCTTA	2580
AAATTCAGTG	CATTCTTGAT	CCACAAGTG	AGTTCCAGGA	CTTACAGAG	AAACCCCTGTG	2640
TTCAAATTTT	TGTGTTCAA	GTACCCCTGG	CTTACAAAGT	GAGTTTCAAG	TCCGATAGGG	2700
CTACACAGAA	AAACCATATC	TCAGAAAAA	AAAAAGTTCC	AAACACACAC	ACACACACAC	2760
ACACACACAA	ACACACACAC	ACACACACAC	ACACACACAC	ACACACACAC	ACACACACAC	2820
AAATGCTGCC	TAAATAAAT	ATTTTCTG	CCAAAGTGAA	AGCAAAATCA	TATGAAGAGG	2880
TACTCTGAGA	AAAAATAAAT	AAAAACGGGC	TTTTTAATCA	TTCAGACAT	GTTTTATATT	2940
AACTCTGAAAT	TTAGTCTTGG	AAAAAGGGGC	GGGTGTGGGT	GAGTGAGGGC	GAGCGAGCAG	3000
ACGGCGGGC	GGGCGGGTGA	GTGGCGGGC	GGGTGTGGGT	GAGTGAGGGC	GAGCGAGCAG	3060
ACCCCAAGCG	GTAGAGTGT	TTAAAAATGA	GACCTAAATG	TGTTGGAACG	GAGGTGCGCG	3120
CCACCTCTCT	CTTCACTGTC	TTAGATGCTC	CCTTCCCTCT	ACTGTGCTCT	CTTCCCTTAA	3180
CTGTGCTTAA	CTGTGCTGTC	TCCCTCACCC	CGCTGATTCG	CCAGCGACGT	ACTTTGACTT	

CAAGAACGAT	TTTGCCCTGT	TTCACCGCTC	CCTGTCATAC	TTTCGTTTTT	GGGTGCCCGA	3240
GTCTAGCCCG	TTGCTATGTT	TCGGGCGGGA	CGATGGGGAC	CGTTTGTGCC	ACTCGGAGGA	3300
AGTGGTGGGT	GGGTACGCTG	CTCCGTCGTT	CGTGCCTGAG	TCGCGGAACG	TGAGCTCCGG	3360
AGACCCCTCC	GAGAGACAGA	ATGAGTGAGT	GAATGTGGCG	GCCTGTGAGC	GATCTGTGAT	3420
GGTTTGTATG	GTTGATCGAG	ACCATTTGTCG	GGCGACACCT	AGTGGTGACA	AGTTTCCGGA	3480
ACCGTCCAGG	CTCTCAAGT	TGCTGACACA	GAGAGAGGAA	GTGCTGTGTG	TGAGGCGACC	3540
AGGGTGACAG	GAGGCGGGCG	AAGCAGGCGG	GAGCGTCTCG	GAGATGTGTT	CGTGTTTAAG	3600
GACGGTCTCT	AACAAGGAG	TCGTACAGGG	AGATGGCCAA	AGCAGACCGA	GTTCGCTGTAC	3660
GCCTTTTGGC	GAAAAATGCT	AGGGTTGTGTG	GCAACGTTTAC	TAGGTCGACC	AGAAGSGCTTA	3720
AGTCTTACCC	CCCCCCCCCT	TTTTTTTTTT	TTTCCTCCAG	AAGCCCTCTC	TTGTCCCCGT	3780
CACCGGGGCG	ACCGTACATC	TGAGGCCGAG	AGGACGCGAT	GGGCGCGGCT	TCCAAGCCCG	3840
TGTGTCTCGG	CCAGCTGGCG	CTTCGGGTCT	TTTTTTTTTT	TTTTTTTTTT	TTTTTCTCTA	3900
GAGGCTTGT	CTGTGCTGTG	CACCGGGGCG	GGTGTACTTC	TGAGGCGCGAG	AGGACGCGAT	3960
GGGCCCCGCG	TTCCAAGCCG	GTGTGCTCG	GCCAGCTGGA	GCTTCGGGCT	TTTTTTTTTT	4020
TTTTTTTTTT	TTTTTTTTCT	CAGAAGCCTT	GTCTGTGCTC	GTCCACGGGG	GCCTGTGACT	4080
TCTGAGCGCG	AGAGGACCGG	ATGGGTTCGG	TTCCAAGCCG	ATGTGGCGGG	GCAGCTGTGA	4140
GCTTCCGGTT	TTTTTTTTTC	CTCCAGAAGC	CCCTCTCTGT	CCCGCTCAC	GGGCGCGCTG	4200
TACTTCTGAG	GCCGAGGAGA	CGTGATGGCG	CCGGTGTCCA	CGCGATGTC	GCCTGCTGAC	4260
CTGGAGCTTT	GGATCTTTTT	TTTTTTTTTT	CTTCAGAAAG	CCCTCTCTGT	TCCTCGCTAC	4320
CGGGGGGACC	TTACATCTGA	GGGCGAGAGG	ACGTGATGGG	TCGGGCTCTC	AAGCCGATGT	4380
GGCGGGGCCA	GCTGGAGCTT	CGGGTTTTTT	TTTTTTCTCT	GATGGAGGCT	CTCTGTCTCC	4440
CGTCACCGGG	GGCGCTGTAC	TTCTGAGGCC	CGAAGGAGCT	GATGGCGCGG	GGTCCAGGCG	4500
GGGTGTCCCG	CGGTCCAGCT	GAGCTTTGGA	TGATTTTTTT	TTTTCCCTTC	AGAAGGCCCTC	4560
TTTTGTCCCC	GTACCCGGGG	GCACCTTACA	TCGTAGGCGG	AGGAGACACG	ATGGGCGCTGT	4620
CTTCCAGCCG	GATGTGGCCC	GGCCACTCG	AGCTTCGGGT	CTTTTTCTTC	TTTGTCTCTC	4680
CAGAAGCCCT	GTCTGTGCTC	GTCAACCGGG	GGCGCTGACT	TCGTAGGCGG	AGAGGACGCG	4740
ATGGGCGCGG	CTTCCAGAGC	GGTGTGGCTC	GGCCAGCTGG	AGCTTCGGGT	ATTTGTTCTT	4800
TTTTTTTTTT	TTCTCCAGA	AACCTTGCTC	GTGCTGTGTA	CCCGGGGCGC	TTGTACTTCT	4860
GATGCCGAGA	GGAACGGAGT	GGCCCGTCTT	GGCCCGGCTA	GGTGGCGGGT	CAGCTGTAGC	4920
TTTGGATCTT	TTTATGCTTT	TTTTCTTCCA	GAAGCCCTCT	CTTGTCCCGG	TCACCGGGGG	4980
CACCTTACAT	CTTAGGCTTA	GAGGACACGA	TGGGCCCCGG	TTCCAGCCCG	TTATGGGCGG	5040
GTACAGTGA	GCTTTGATCT	TTTTTTTTTT	TTTTTCTTCA	GAAGCCCTCT	TGTCGCCGCT	5100
ACCCTGGGCA	CTGTACATCT	GAGGCGGAGA	GGACATTATG	GGCCCGGCTT	CCAATCTCGAT	5160
GTGGCCCGGT	CAGCTGGAGC	TTTGGATCTT	ATTTTTTTTT	TAATTTTTTC	TTCCAGAAGC	5220
CTCTCTGTTC	GCTTACCCGG	TGGCACGGTA	CATCTGAGGC	CGAGAGGACA	TTATGGGCGC	5280
GGCTTCCAGG	CCGATGTGCG	CCGCTCAGCT	GGAGCTTTGG	ATCTTTTTTT	TTTTTTTTCT	5340
TTTTTCTCTC	TGAGGCCCTC	TCGTCCCTGT	TCACCGGGGG	CCCTGTACGT	TTGAGGCGGA	5400
GGGAAAGCTA	TGGCGCGGCT	TTTCTTTTAT	TGACCTGTGC	GTCTTATCAG	TTCTCCGGGT	5460
TGTCAAGGTC	GACCACTTGT	TCCTTTGAGG	TCGGTTCTTT	TTCTGTTATG	GCTCATTTTT	5520
GGCCCACTCT	CCCGAGTATG	ACTTCCAGGC	GTGCTGTGCT	GCCTGTCACT	TTCTTCCCTG	5580
TGCTTTTAT	CTTGTGATC	TTTTTATCT	GTTCCTTATG	GACCTGTGCT	TAGGTACTGA	5640
CACGCTGTCC	TTTCCCTATT	AACACTAAAG	GACACTATAA	AGAGACCCTT	TGATTTTAAG	5700
GCTGTTTTGC	TGTGCCAGCC	TATCTTTTTT	ACTGGCTTGG	GTCTGTCCGG	GTGGCTGAAG	5760
CTGTCCCCGA	GCCACGCTTC	CTGCTTTCCC	GGGCTTGCTG	TTTCCGTGTG	CTTGTCTTGG	5820
GCAGCTTGTG	ACAACCTGGC	CGTGTGACTT	TGCTGGCTGT	CAGACGTGTT	TTCCGATTTC	5880
CCCGAGGTGT	CGTTGTACAC	CCTGTCCCGG	TTGGAATGCT	GGAGCCAGCT	GTGGTTGAGG	5940
GCCACTTAT	TTCCGGTCTT	TTTTTTTTTT	TTTGAATGCT	TGGAGTCCCG	GAACTCCCGC	6000
CTTTTCTCT	TTCCGGTCTT	TCTTCCACAT	GCCTCCCGAG	TGCAATTTCT	TTTGTTTTTT	6060
TTCTTTTTTT	TTTTTTTTTT	TTTCTTTTTT	GGAGAGTCCC	GAGTACTTCA	CTTGTGTCTG	6120
TGGTGTCCAA	GTGTTTCATG	CACGTGCCCT	CCGAGTGAC	TTTTTTTTTT	GGCAGTCTCT	6180
CGTTGTGCTC	CTTGTGTCTG	TGCTGCCCGG	TATCAGTAAC	TGTCTTGGCG	CGCGTGTGAAG	6240
ACATTCTTAT	CTCGCTTGTT	TCTCCCGATT	GCCTGCTGTT	GCTCACTCTT	AGATCGATGT	6300
GGTGTCCCGG	AGTTTCTCT	GGGCGAGGGC	CAGCGCGCGC	CAGGCGAGGG	ACGACGATC	6360
ATGGCGAATG	GCAGCGCGCT	TTCTCTTTCT	GCACGCGGCG	CTTGTGCTCT	CGACCCATC	6420
CGTCTGCCGG	CTGTGTGTGG	AAGGCAAGGG	TGCGGCTCTC	CGCCCGGCGC	CTGCCCGCGG	6480
CGCATCTTGT	TCAGTGTGTT	GCCTGTGCTC	TGTGATGTGT	TTGAGTGTGG	GGTGTGTGCC	6540
TCACGTGTTT	CACTTTGCTC	GTGCTCGCTG	TGACCATGTT	CCGAGATGTC	TGGATGTGCG	6600
CCGGTGGCGT	TGCATACCTT	TCCTGCTCTG	TGTCTGACAG	CGCTGTGTTT	TGTAAGCGTC	6660
GAGGTGCTTC	TGGAGCGTTC	CAGGTTTGTG	TCTTCCGCTG	CTGCTTCTGA	CTGCTGTGTT	6720
CGGCTCCCCA	TTCCCTGGTG	GGGCTCCGTT	CGCTCGCTGC	CGTGTGTGCC	TTGAGGCTTTG	6780
TGCTGTAGAA	GTCCTGTGAG	TCGCGGTGAG	GTACAGAAAG	AGGCGCAAGA	CCCCCTTCTT	6840
TGCTCGGGTG	AGGCGCCACC	CCCGCACTA	GTACGCTGCT	CGCTAAGGCT	GGTGTCTGAGC	6900
GGTCGCGGCT	GGGTTTGGAA	AGTTTCTTCA	GAGACTCAIT	CGTTTCCCGT	GGGAGCTTCT	6960
GAGAGCGCTG	GCTTTCGGGG	GGGACCGGTT	CAGAGGCTCT	CCCTGTGCGG	GGGAGCTCTAG	7020
AATGCCCTTG	GAGAGAAACC	TTCTCTGTTC	CGCAGACCCC	CCCGCGCGGT	CGCCCGCGTG	7080

TTGGTCTTCT	GGTTTCCCTG	TGTGCTCGTC	GCATGCATCC	TCTCTCGGTG	GCCGGGGCTC	7140
GTGCGGGTTT	TGGGTCCGTC	CGGCCCTCAG	TGAGAAAGTT	TCTCTCTCTA	GCTATCTTCC	7200
GGAAAGGGTG	CGGGCTTCTT	ACGGTCTCGA	GGGGTCTCTC	CCGAATGGTC	CCCTTGGAGG	7260
CTCGCCCCCT	GACCGGCTCT	CGCGCGCGCA	GCCTTTGCTC	TCTCGTCTAC	CTCGGGCCGC	7320
GGCCTCCCCG	CTCCGAGTTT	GGGGAGGGAT	CACGCGGGGC	AGAGCTGTCT	TGTCTCTCTG	7380
CCCTGTCTCG	GGAGCATGTG	GCTCGGCTTG	TGTGTTTGGT	GAGCTGGGAG	ACGGCTCCGT	7440
GCACACCCCC	CGGTGCGGCT	ACTTTCTCTC	CTCTCTGAGT	CGCCGCGTGT	GGACCGGGTG	7500
TGGGTAGGCG	ACGGTGGGCT	CCCGGGTCCC	CACCGCTCTT	CCCTTGCCTC	ACCGCTGGCT	7560
CTGTTCGCGT	GGTCTCCCTC	CGCTCGCGTC	CACGACTTTG	CGCGTCCCG	CGACGGCGGC	7620
CTCGCGCGCG	CGTGTGTGCT	TCTCGGCTCT	TCTCGGCTCT	TGTGTGTGTG	TCGCTCTCGC	7680
CCCCCTTTCC	GAGAGGGTCC	GTGCTCGGCG	TTGAGAAATC	GGGAGATGCC	GCTTCCCTCT	7740
CTCTCGGGGT	AGAACGGGCT	TTGGCCGGCT	CGCGCGCGAG	TGCTCTCGG	GGACGGAGCC	7800
GTTCTGTGGG	GGTCTTCTGC	GTGAGGCATC	GGTGTGTGTT	CTCTGACCTG	GGGACCACTT	7860
CGCGCTCGGG	TCTCTCGGCT	CCCGGGGGGC	CGTGTGTGTT	CATCGGTCTC	TCTCTCTGTG	7920
CGGTGTGCTG	GACTGCTCAG	GGGAGTGGTG	CAGTGTGATT	CCCGCGGCTT	TTGCTCTCGC	7980
TGCGCTGACG	GGTCCGACGC	CCGAGCGGTC	TCTCGGTCCC	TGTTGAGGAC	CCCTTCCCGG	8040
GAGGGGGCCG	GGTCCGCGCG	CTTTCGCGTC	GTGCGCGGCC	CTGCTTCTCG	TGTGTCTGTG	8100
CCCCCTCCCC	GTCGCGCGCA	CGCGTCTTCT	TTTCTCTCTC	CCGCCCTCTC	CTCTGAGACT	8220
ACCCGTGGCC	GGTCTGTGCG	ACCCCGCCGA	TGGGGCGCGC	CGGGCACGTA	CGCGTCCGGG	8280
CGGTTCACCG	GTCTTTGGGG	GGGGCGCGAG	GGTAAAGAAA	CTCGGTCCGG	CGGGCGGGAG	8340
GAGCTGGGTA	TTGAGGGGCG	TCCCGGCCCC	CGCGCCGTGG	CGGTGTCTTG	CGCGGCTCTG	8400
GAGAGGGCTG	CGTGGGAGGG	GAAAGGTTTG	CCCGCGGAGG	GCAAAGGGAA	GAGAGCTAGC	8460
ACTGTGCTATT	GTCCCGACGG	TGTGGTGGTC	TGTTGGCGCA	GGTGGCTCTG	GGGGGCTCGT	8520
CGCGCCCTGT	CGCGCTTCGG	GAAAGCGCGT	GTGGGGGCTT	CGCGGAGTGC	GAGAGTGGGT	8580
ACCTCTGGCG	TGGGATTAAC	CCCGCGCGCG	TGTCGCCGTG	TGGCGGTGGG	GGCTCCGGTC	8640
GATGTCTACG	CTGCTTCTCC	CGAGGTTCTA	GGCTTCTCTC	GCGCGGGGCTC	TGCGGCCCTCC	8700
CTCGTGTCTT	CGCTTCTCGG	GGGTCTCAAG	CGCTCGTGGT	CTCCCTCTCC	TCGCTCTCTC	8760
CATCTCTCTG	GCAATGGGCT	CGCCCGAGTT	CACGGTGGGT	TGCTTCTCCG	CTTCCGCTCTC	8820
TGCGCGGGGG	CTGCGCGCTG	TCCGGTCTCT	CTTGCCTGAC	CCCGGTGGGC	GTGGTCTTCT	8880
CTCGCGCGCT	TCGCGGAGTC	GTCGCTTCGC	CCGAGGGGTC	AGGGGGCTTC	CCGGTTCGCC	8940
GACGTTGCGC	CTCGCTGCTG	TGTGCTTGGG	GGGGCGCCGC	TGCGCGCTCT	GCCCGCCGCT	9000
GAGCCCTCTG	GCGACCCGCG	GGTGTGCGGT	TTGCGCGCGC	GGTCAAGTGG	GCCTTGGCGT	9060
TGTGTTCGCT	CGGGAGCGGT	TCCGCTTCGC	GGCGGCTAGA	CGCGGCTGTC	GCGGGCTCCG	9120
CACGGGTGGC	CTATCCAGGG	TGTCGCCCGC	CGGACCCCGC	CTTGGCCGTC	CCGGTGTGGG	9180
TCGTTTGGTG	GGGGAGTGA	TGGTGTACAC	GGTCAITCCC	TCCCGGTGGT	TTTGAAGTTC	9240
TCGCGCGTGT	TCGCGGCTTC	TTTCCGCCAA	CCCCACGCG	AACCCACAC	CTGCTCTCTC	9300
CGCGCCGCTG	CGGTGCAAGT	TCCGGCTCTC	CCGATGCGCA	GGGGTTCGGG	ATTGTGCGCG	9360
GGGACGGAGG	GGAGAGCGGG	TAAGAGAGGT	GTGCGAGAGC	TGTCGCGGGG	GACGCTCGGG	9420
GTGTCGTTTG	CGCGTTCGCT	GTGCTCGCGG	ACGGGTTTGT	TGCGAACCCG	ACGGGCTCGG	9480
TCGCGCGCGA	TGCATCTCTC	CGTTTCGCGC	GAGCGCCCGC	CGGCTTCACC	CCGGGTTTGT	9540
CTTCCCGCGA	GGCTCTCCGC	CGCCCGCGCC	TCTCTCTCTC	CTCTCGGCTC	CTCTGTCCCG	9600
CTTGGTCTGT	TCCCAACCCC	GACGCTCCGC	TGCGGCTTCT	TTACCTTGTG	GATCTCGGCA	9660
GGTAGCATAT	GCTTGTCTCA	AAGATTAAAG	CATGATGTC	TAAGTACGCA	CGCGCGGTAC	9720
AGTGAACACT	CGAATGGCTC	ATTAAATCAG	TTTGTGTCTC	TTTGTGTCTC	CGCTCTCTCT	9780
CTACTTTGAT	AACTGTGTGA	ATTCTAGAGC	TAATACATGC	CGACGCGGCG	TSACCTGCGT	9840
TCCCGGGGGG	GGATGCGTGC	ATTATCAGA	CTAAAACAGC	CCCGGTGAGC	TCCTCTCCGG	9900
CTCCGCGGGG	GGGTGCGGCG	CGCGGCGGCT	GGTGACTCTA	GATAAACCTG	GCGCGATGCG	9960
ACGCCCCCCG	TGGCGGGGAC	GACCCATTCC	ACCGGCGAAT	CTATCAACTT	TGATAGGTAG	10020
TCGCGGTGAC	TACCATGTGT	ACCACGGGTC	ACGGGCGAGT	AGGCTTCAAT	TCGGAAGAGG	10080
GAGGCTGAGA	AACGCTTACC	ACATCCAAAG	AGGACAGGAG	CGCGCGAAAT	TACCCACTAT	10140
GACCCCGGGG	AGGATGTGAC	GAAAAATAAC	AATACAGGAG	CTTTTGGAGT	CCCTGTAAAT	10200
GGAAATGAGTC	CACTTTAAAT	CGATTAAACG	GGATCCATTG	GAGGGCAAGT	CTGGTGCCAG	10260
CAGCGCGGCT	AATTTCCAGT	CGATAGCGCT	ATATTAAAGT	TGCTGCGAGT	AAAAAGCTCG	10320
TAGTTGGATC	GGTTCGAGTC	CGCGCGCGCT	CGCGCGAGG	CGAGTCAACG	CCGCTCCCGG	10380
CCCTTGTGCT	CTCGCGCGCC	CCTCGATGCT	CTTAGCTGAG	TGTCCCGCGG	GGCCCGGAAG	10440
GTTTACTTGT	GAATAATTAG	AGTGTTCAAA	CGAGGCGGCA	GCGCGCTGGA	TACCGAGCAT	10500
AGGAATATAT	GAGTAGAGAC	CGGGTCTTAT	TTTGTTCGTT	TTCGGAATCT	AGGCATGATG	10560
TAAGAGGCGAC	CGCCCGGGGC	ATTCTGATTT	CGCGCTAGA	GGTGAAATTC	TTGGACCCGC	10620
GCAAGACCGA	CCAGAGCGAA	AGCATTTGCC	AAGAATGTTT	TCATTAACTA	AGACGAAAG	10680
TCGGAGGTTT	GAAGACATAT	AGATACCGCT	GTAGTTCCGA	CCATAAACGA	TGCCGACTGG	10740
CGATCGCGCG	GGCTTATCTC	CATGACCCGC	CGGGCAGCTT	CCGGGAAACC	AAAGTCTTTG	10800
GGTTCGCGGG	GGAGTATGCT	TGCAAAAGCT	AAACTTAAAG	GAATTGACGG	AAGGGACACA	10860
CCAGGATGGG	GCGCTCGGCT	TAATTGACT	CAACACGGGA	AACCTCACCC	GCCCGGACCA	10920
CGGACAGGAT	TGACAGATTG	ATAGCTCTTT	CTCGATTCCG	TGGGTGTGTG	TGCATGGCCG	10980

TTCTTAGTGTG	GTGGAGCGAT	TTGCTGGTT	AATTCGGATA	ACGAACGAGA	CTCTGGCATG	11040
CTAACTAGT	ACGCGACCCC	CGAGCGGTG	CGCTCCCCA	ACTTCTTAGA	GGGACAAGT	11100
GCCTTCAGCG	ACCCGAGATT	GAGCAATAAC	AGGTCTGTGA	TGCCCTTAGA	TGTCGGGGG	11160
TGCACGCGCG	CTACACTGAC	TGGCTCAGCG	TGTGCTTACC	CTGCCCGCG	AGGCGCGGGT	11220
AACCCGTTGA	ACCCCAATTCG	TGATGGGGAT	CGGGGATTGC	AATTAATTC	CATGAACGAG	11280
GAATTCCTCA	TAAGTCCGCG	TCATAAGCTT	CGCTTGATTA	AGTCCCTTCC	CTTTGTACAC	11340
ACCGCGCGCT	GCTACTACCG	ATTGATGGT	TTAGTGAGGC	CCTCGGATCG	GCCCGCGCGG	11400
GGTCGCGCCA	CGCCCTGCG	GGAGCGCTGA	GAAGACGGTC	CGCTTGACT	ATCTAGAGGA	11460
AGTAAAGT	GTAAACAAGT	TTCCGTAGGT	GAACCTGCG	AAGGATCACT	AAACCGGGAGA	11520
CTGTGGAGA	CGGCGCGGT	GGCCCGCTCT	CCCGTCTTG	TGTTGTGTTC	CGCGGCGAGG	11580
CGCTGCGCTG	CCGCGTCCCG	TCGCCCGCTG	GTGGAGCGAG	GTGTCTGGAG	TGAGGTGAGA	11640
GAAGGGGTGG	GTGGGTCGCG	CTCGGCTCCG	TCGTGGACCG	CCTCGATCTT	CCCTCCCCC	11700
TCCCTCTCC	CTCGTCCGCG	TCGACCTCG	CCACCTTACC	CGCGGCGCGG	CTGCTCCGCG	11760
GGTCTTGCC	TCITTCGCT	CGGCTCTTTC	CGTGTCTACG	AGGGGCGTGA	CGTGGTTACG	11820
GGTTTTTGA	CCGTCCCGCG	GGCTTCTCGT	CGTGGGCGG	CGCGCTTTGC	CTCTCCGCA	11880
CCCATCCCG	CGCGGCTCT	GGCTTTTCTA	CGTGTGCTGG	GGCGGTTGTC	GGCTGTGGG	11940
GGATGTGATG	GTGCGGTGTG	GGCTCGCCCG	TCCCGATGCC	ACGCTTTTCT	GGCTCTCGCT	12000
GTCTCCCG	CTCTGCTCC	GGGTACCTAG	CTGTGCGCTG	CGCGCGCGGA	GTCTTAAGA	12060
CCCGCGGGG	GTGCGCTGAC	CGCCCCGAG	GTGCGGGGCG	GGTGGGCGC	GTAGCGAAGT	12120
CGGTCTTGC	GGCGGCTCTC	CTCAGACTC	CATGACCTC	CTCCCGCGC	TGCGCCGCTT	12180
CCGAGAGCG	CGGTCTGTGT	GGGGGTGGA	TGCTTGAGC	CGCTCGGCG	GGCGGTGGG	12240
CCGAGACCG	CGCGCGGCT	TGCTTTTCCG	CCCGGCTG	CTCTCTGCTG	TGCGGCTCTG	12300
GGGTTCGCT	GTCTGTCCCG	TGTTTCTCCG	CTCCGAGCCC	TTTCTTTTTC	CTCCCCCA	12360
CACGTGTCTC	GTCTGTCTCC	TGCTGTGCTG	CTCGGCGCG	CCCTCTGCTC	GTCTGTCTC	12420
CTCTCTCTC	CGGGGAGAG	AGGGCGCGG	CTCGGAGCTA	CGCTGTGATC	ACTCGGCTG	12480
CGTGAAGTTC	CTACACCG	AAATCCGAT	ACGACTTTTA	CGGCTGATC	ACATCATGAT	12540
TGCGTGTGAT	AAGAACCGAG	CTAGCTGCGA	GAATTAAGT	CGGCGCTACG	CCTGTCTGAG	12600
CATGACACT	CGATACGAC	TTGCGGCCCC	GGGTTCCTCC	CGGGCTACG	CTGTCTGAG	12660
CGTGCCTTGA	CGATCCCGA	CGTACCCGCG	GTGCGTGGT	GCTGCGCGGT	TGGGAGTTTG	12720
CTCGCAGGG	CTGCTCCCA	ACCCGAGTCT	GGCCCTCTCG	CTCCGGAAGT	TACAGACGTG	12780
GGCGGTTGT	CGGTGTGGG	CGCGCGCCCG	CGTCTGCTCC	CGCGGCTC	CCCGCGCAT	12840
CGCGCTGCG	GCTTCTTCC	GCTCCGCGCT	TCCCGCCCTC	GGCGGTGAC	CCCGGCTCTG	12900
CGCTCGGCT	CGCGGCTCC	GGACCGCTGC	CTCACAGTC	TTTCTGCTC	CGGTGCCCG	12960
TGGGAACCA	CGCGCCCCC	GTGCGCCCG	GGGGTGGCG	CGTCCGATC	TGCTCTGCT	13020
GAGGTGTGCG	TTGAGGGTG	TGCTGCGCC	GAGGTGGTGG	TGCGTCCCT	CGCGCGCGG	13080
GTGTGTGCG	GTGCGGTG	ACGAGGGCG	GTGCTGCGC	TGCGGTGTTT	GTCTGTGTGT	13140
GTGTTGCTCT	TCGCTGGGG	GAGCGGGGT	CGACCGCTC	CGGGGTGTC	CGGCTCGCC	13200
GGCGCGCGC	ACCCCTCGCG	TTGTGTGAG	GGAGAGCGAG	GGCGAGAACG	GAGAGAGGTG	13260
GTATCCCCG	TCGCGTGGG	AGGAGGGT	TGGCGTCCG	CGTCCGACG	TCCTTCTCT	13320
CGTGTGGG	CGCTCTCGC	CCGACGCGG	CGCTAGGGG	CGGTGCGGCG	CGGTGGCCCT	13380
CGTGTCTCT	CTTCTCTCT	GCTTCTCTT	ACCCGCGCG	GTACCCGCTC	CGCGCCGCG	13440
CCGCGGGAG	CCGCGGCTC	CGTGGCGCA	TGCGAGTCA	CCCCGGGTG	TGCGAGTTCT	13500
GGGAGGGAG	GGGCTCTG	GACCCGTTC	GTCCGCGGT	CCTCGGGGG	GACCCGCGGT	13560
CTGTGGGCT	TGCTCTCCG	GGGTGCGTG	TGAGTAAGAT	CCTCCACCC	CGCGCCCTC	13620
CCCTCCGCG	GGCCTCTCG	GGACCCCTG	AGACGGTTCT	CGGCTCTCT	CTTCTGAGC	13680
GGCGGCTGCG	GTCTCTTTC	CGCCCGCTC	CTCGCTCTC	TCTTCCCGCG	GTGGGCGCG	13740
TGTCCTCCCT	TTCTGACCG	GACCTCAGAT	CAGACGTGCT	GACCGGCTGA	ATTAAAGCAT	13800
ATTAGTCAG	GGAGGAAAG	AAACTAACCA	GGATTCTCTC	AGTAAACGCG	AGTGAACAGT	13860
GAAGAGCCCA	CGCGCGATC	CCGCGCGCG	GGGAAAGTCT	CGGGAATGTC	GGCTAGCGAA	13920
GACCACTCTC	CGCGCGCGC	TCGTGGGGG	CCCAAGTCTC	TCTGATCGAG	CGCCAGCCCG	13980
TGAGACGGTG	GAGCGCGTGA	CGCGCCCGG	CGCGCGGCT	TCGGGTTCTC	CGCGAGTGG	14040
GTGTGCTTGG	AATGCAAGC	AAAAGCGGTG	GTAACCTCCA	TCTAAGGCTA	AATACAGGAT	14100
CGAGACCGAT	AGTCAACAG	TACCGTAAAG	GAAGAGTTGA	AAGAAGCTTG	AACAGAGAGT	14160
TCAGAGAGCG	GTGAAACCGT	TAAAGAGTAA	ACGGGTGGGG	TCCGCGAGT	CGCCGCGAG	14220
GATTCAACCC	GGCGCGCGC	GTCCGCGGT	CCCGGTTGTT	CCCGGTTGTT	CTTCCCGCT	14280
CCCGCTTCTC	CCGAGACCTC	CCACCGCGC	GTGCTTCCCG	TCTTCTCTCC	CTTCCCGCT	14340
GCCTCCGCG	CGGGCGCGG	GGGTGTGTGT	GGTGTGGCG	CGCGGCGCG	CGCGGCGGTG	14400
GGGTCCGCGG	GGGACCGCGC	CGCGCGCGC	ACCGCGCGC	CGCGCGCGCA	CTTCCACGCT	14460
GGCGGTGCG	CGGACCGCGC	TCGGGAGCG	CCGGAAGGCG	CGCGTGGGA	AGGTGGGCTC	14520
GGGGGGGCG	CGGCTCTCAG	GGCGCGCGA	ACGACCTCAC	CCGAGTTGAT	AGCGCTTCC	14580
GGCGCGGCTT	GTCCGCGCTC	CGCGCGCGA	GGAGGCGAGA	TACCGGTGCG	CGGCTCTCC	14640
CTCTCCCCCT	GTCCGCGCTC	CGCGCGCGC	TGGGGTGGG	GGCGCGGCG	CCCTCCAC	14700
GGCGGACCG	CTTCTCAC	CCCTTCTG	GGCTCTCTC	GGCGCGGCG	GGGGGCGGG	14760
CGCATGCGCT	CGCATGCGC	CGCGCGCTC	TCGCGCGCT	GGGTCCGCG	GGGACGCTG	14820
GTCACTGCTC	TCCGACGAA	CGCGAGCGCA	CGGGTCTGCG	GGCGATGTC	GCTACCCACC	14880

CGACCCGTCT TGAAACACGG ACCAAGGAGT CTAACGCGTG CGCGAGTACG GGGCTCGTCC 14940
 GAAAGCCGCC GTGGCGCAAT GAAGGTGAAG GGCCCGCGCC GGGGGCCCCG GGTGGGATCC 15000
 CGAGGCCCTC CCAAGTCCGC GAGGGCGCAC CACCCGGCCCG TCTCGCCCGC CGCGCCGGG 15060
 AGGTGGAGAC CGAGCGTACG CGTTAGGACC CGAAAGATGG TGAACATATG TGGSCGAGG 15120
 CGAAGGCCAGA GGAACACTCTG GTGGAGGTCC GTAGCGGTCC TGACGTGCAG CCGTCTCGTC 15180
 CGACCTGGGT ATAGGGGCGA AAGACTAATC GAACCATCTA GTAGCTGTGT TCGTCCGAAG 15240
 TTTCCCTCAG GATAGCTGCG GTCTCTCGTC CGACGTATCG TACTTTATCT CGTAAAGCG 15300
 AATGATTAGA GGTCTTGGGG CCGAAACGAT CTCAACTATC TCTTCAACTT TAAATGGGTA 15360
 AGAAGCCCGG CTCGCTGGCG TGGAGCCGGG CGTGGAAATG GAGTGACCTT TGGGCGACTT 15420
 TTGTGAATGA GAACTGGGCG TCGCGGATGA ACCGAACCGC GTTGAATAGG GCGCGATGCC 15480
 GACCTCTAGC AGACCCCGCA AAGGTGGTGT GTTGATATAG ACAGCAGAGC GGTGGCCATG 15540
 GAAGTCGATA TCCGCTAAGG ATGTGTATAA AACTCACCTG CCGATCAAC TAGCCCTGAA 15600
 AATGGATGGC GCTGGAGCGT CGGGCCCATC CCGCGCCGTG CGCGAGTCTG GAACGGAACG 15660
 GGACGGGAGC GGCCGCGGT GCGGCTCTCT CGGGGTCTGG CCGGCTCGGG GTTGCCTGCG 15720
 CCCC CGCCTT CCGCTCGCGC GCGCGGTTCG CCGCCGCGGC GTCCGCCCCC CGGAGGACTA 15780
 CGCCGCGAGC AGTAGGAGGG CCGCTGCGGT CCGCTGTGTA GAGCCTTGAA GCGTAAAGCG 15840
 TGGAGCCGCG CAGAGTGGAG ATCTTGGTGG TCCATGTGAA TATTCAAACG AGAATTTGA 15900
 AGAGCCGAAG GGAGAAAGGT TCCATGTGAA GCGCATGAGC TCTCGTGCC CTCGCGCTGT 15960
 CGAAAGGGAG TCGGGTTCAG ATCCGGAAT CCGGATGGCG GGAGATGGGC CGCGCGAGCG 16020
 CAGTGGCGGT ACSCGACCGA ACSCGACGAG GCGCTCGGGA AGCTTCGGG AGATTTCTCT 16080
 TTTCTTTGTT TTGGAAGCG AGGCGCAGG TCCCGAGATA CGCGCGCGGA GAGTCTTCTT 16140
 GGGGAGAGGG TCGCGTCTTC GCGCGGTCTC GGTGAGCTCG CCGAGAGAGG GCGCGTGGC 16200
 ACTTCCGAGT TCTAATCTG CCGCCGCGCC GTTACCATAT CCGCAGCGCT TGAATATCC 16260
 CTGGCGCGCG CGCCGCGCTT GAACAATGTA GTTAAAGGAA TCCGCAAGG CTGATCCGTA 16320
 CCGCTCTCTT CGCCCGCGGC CGCCCTCTCC GTTCTCCCG CCGGCGCCCG CTGCTCCCCG 16380
 CGTCTCGTCC ACTCTCTCTT CCCCCTCTT CTTCCGTCG GCGCGCGGCT GCGCGCGGTA 16440
 CGCGCGCGCG CGGCTCTCGG GCGCGCGGTC CAACCCCGCG GGGGTTCGCG AGCGGGAGCG 16500
 ACCAGCGGTC CCGGCTGGGG GCGGGGCGCC GGACACTCGG GGGCGCGGCG CGCGCGCGGA 16560
 CTCTCGAGCG GAGCCCGCGCC CTTCGCTGAG ATCGCCTCAG CTGCGCGGCG CTGCGCGGCC 16620
 GGTCCCGCGG AGCCCGCGCG GTGCGCGCG GGGTCCCTC CCGCGCGGCG CCGCTCCAC 16680
 CCCCCATCGT CCGTCTCCGA GGTGCGTGGC GGGCGCGGCG GGGCGTGTCT CGCGCTGTG 16740
 GGGGGAACCT CGCGGTTCGG GTTCCCGCG CAGCCGCTAG CCGCGCGGCG CGGCTTTCCG 16800
 CGCGCGCGCT CCGCTCTCGG CGCGCGCTAG CCGGAAGGCC CGCGCGGCT GTTGACCGTA 16860
 GAATCGCACT GTTTAATTA AACAAGCAT TCTCTTAAG TGTCCATAC TACTATCCAG CGAAGACCA 16920
 TGTGATTCTT GCCCAGTGCT CTGAATGTCA AAGTGAAGAA ATTCATATGA GCGCGAGTAA 16980
 ACGGCGGGAG TAACATGTAG TCTCTTAAG TAGCCAAATG CCGCGCGGT GTTGACCGTA 17040
 CGCGCATGAA TGGATGAAG AGATTCCAC TGTCTTACC TACTATCCAG CGAAGACCA 17100
 GCCAAGGGAA CCGGCTTGGC GGAATCAGCG GGAAGAAGAG ACCCTGTGTA GTTGAATCT 17160
 AGTCTGGCAC GGTGAAGAGA CATGAGAGT GTAGAATAAG TGGGAGGCC CCGCGCGGCT 17220
 GCGCCGTCTT CCGCTCGGGG TCGGGCGCAG CCGGCTCTCG GGGCGCGCG TGAATATCCA 17280
 CTACTCTCAT CGTTTTTICA CTGACCCGCT GAGGCGGGGG GGGCGAGCCC GAGGGGCTCT 17340
 CGCTTCTGCG GCCAAGCGTC CGTCCCGCG GTGCGGGCG GCGCGAGCCC CGCGCGGAG 17400
 AGTGCAGGCT GGGGAGTTG ATCGGGGCGG TACACTGTCT AAACCGTAC CAGAGTGTCT 17460
 TAAGCGAGTC TCAGGGAGGA CAGAAAACCT CGGTGAGGA GAAGGGAGCA AGCTCGCTCT 17520
 ATCTTGATT TCAAGTACGAA TACAGACCGT GAAAGCGGG GTTACACAG GGTACACTG 17580
 TTGGGTTTTT AAGCAGGAGG TGTCAAGAAA GTTATCCAG GGTATCTCT CTATCATTT 17640
 CCAAGCGTTC ATAGCGAGCT CGCTTTTIGA TCCTTCGATG TCGGCTCTC TGTGTACCT 17700
 GAGGACAGAT TCACCAAGCG TTGATGTGTT CACCACATTA CCGCATATG TAGGTAAGT 17760
 TAGACCGTGT AGAGACAGGT TGTGTTTACC TACTATCTAG TGTATGTGT TGTGATGAG 17820
 CTTGCTCAGT CCGAGGAGAA TCGCGAGTGA TATATCTAGA ACGCCCTCAA TGTCAAGAT 17880
 GCCAATGGGG CGAAGCTACC ATCTGTGGGA TATCTGTGGA GACGCTCTCG TGGCCCGGT 17940
 AACCATACCG AGCGATACCG CAGCGCGGAA CCGTCCCGCG CGCGCGCGCG GGTCTCTCC 18000
 CCGCGTCTCGT CCGCTCTCGG CGGTCTCGGC GTCTCGGTTC GAGAGAGCGT TCGTCTTGG 18060
 CCGCGCGGCG GGGGAGCGCG GCTCTCGGCG CCGCTCTGAA CGCATCTGCG GTTGCGGCT 18120
 CCGCGCTAAG ATTGGTAGA GACACTGCTT CTGGGTGCGG GTTTCGTAG TAGCAGGCA 18180
 GCTCCCTCGT TCGAGTCTAT TGAAAGTCAG CCTCTGCAC AAGGGTTTTT CTCTCGGGG 18240
 TTTCCCTCTG CAGCGCGGCT CCGTCCGAC CGACGTGTC GCGCGCGGCG CGTCAAGGCG 18300
 CGCTCGCTCT GCGCCCGCG CCGGCTTCCG GAACGACCG GTGTGTGTTT GGGGCGGGAT 18360
 CGTCTCTCC TCCGCTCTCC GAGGACGGT CTTCTCTCT TCCCTCTCG TCGCTCTCT 18420
 TGGGTGTGG AGCTCTGTC GTCTCGGACC CCGCGCTGCC GTGCGCTGCC GCGCGAGCC 18480
 CTTGCCCTCC GGCCTTGGCC AAGCCGAGG GCGGAGGAG GCGGAGGAG GCGCGCGCA 18540


```

CCGCGGCGCG GTGACGACG GTGGGATCCC CATCTCGGC GCGTCCGTCG GGGACGGCGG 18840
TTTGGAGGGG CGGGAGGGGT TTTTCCCGTG AACGCGCGCT TCGGGCGCAG GCCTCTGGCG 18900
GCCGGGGGGG CGCTCTCTCC GCCCGAGCAT CCCACTCCC GCCCTCTCTC TTCGCGGCGC 18960
CGCGGCGGGA CGTGGCTAGC AGGGGAGGAT GTGCGCGTGT GGAGGCGGAG AGGGTCCGGC 19020
CGGGCGCCTC TTCCATTTTT TCCCCCCCAA CTTGCGAGGT CGACGAGTAC TCCGGGCGAC 19080
ACTTTGTTTT TTTTTTTTCC CCGGATGCTG GAGTTCGACC AGATGTCGGA AAGTGTCCCC 19140
CCCCCCCCCC CCCCCCGGCG CGGAGCGGCG GGGCCACTCT GGACTCTTTT TTTTTTTTTT 19200
TTTTTTTTTT TTAATTTCCT GGAACCTTTA GGTGACCAAG TTGTCCGCTT TTTACTCCCT 19260
CATATAGGTC GACCAAGTACT CCGGGTGGTA CTTTGTCTTT TTGTGAAATG TCCGAGGTTT 19320
GACCAGATAT CCGAAGATCC TCTCTTTCCC TTACTCTTTC CCACAGGAGA TTTCTCTGTC 19380
TTTTTTTTTT TTTGGTGTGC TCTTTTGTGA CTTATATACA TGTAATATAG GTGTACGTTT 19440
ATATACTTAT AGGAGGAGGT GCACCAAGTAC ACAGAGTATC CTTGGGCGAC ACTTTTGTTT 19500
TCCACCGATG ATGGAGGTCG ACCAGATGTC CGAAAGTGTG CCGTCCCCCC CTCTCCCCCC 19560
CGCGGCTGCG CGGGCTCTAC TCTGGACTCT TTTTTTTTTT TTTTAAATTC 19620
TGGAATGCTA AGGTGCACCA GTTGTCCGTC TTTCACTCAT TCATATAGGT CGACCGGTGC 19680
TGACTTTGCT TTTTCTGAAA ATCCGAGAGG TCGACAGAGT CTCGAGGTCG 19740
ATAAATATAT TGATCTAGT TTTTCTTCTT GTTCTTCAGT TTTGTTTGTG TTTGTTTGT 19800
TTTGTGTTCT TTTGTTTGT TTAGTTTTGT TTGTTTGTG TTTGTTTGTG TTTGTTTGT 19860
TTTGTGTTGT TTTGTTTGT GTTGTGTTGT GTTGGGTTGG GTTGGGTTGG GTTGGGTTGG 19920
TTTGTGTTGT TTTGTTTGT GTTGTGTTGT TTTGTTTGT TTTGTTTGT TTTGTTTGT 19980
TTTGTGTTGT TTTGTTTGT TTTGCGGGT CGAACAGTGT TCCCTAACCG AGTTTTTTTG 20040
TACACAAACA TGCACTTTTT TTAATAAATA TTTTAAAT AAATGCGAAA ATGACCAAT 20100
TATCCCTTTC TCTCTCTCTC TTTTTTAAAT ATTTCTTTG TGTGTGTGTG TGTGTGTGTG 20160
TGTGTGTGTG TCGGTGTGTG TGTGTGTGTG CGTGCAGCGT GCBCGCGCTC GTTTTATAAA 20220
TACTTATAAT AATAGTCCGC CGGGTGTGTG TAGCTTCCG CACTCCAGAG CGAGAGGAGG 20280
CGACACTTCT GAGTTCGAGG CCAGCCTGGT CTACAGAGGA ACCCTGTCTC GAAAATGAA 20340
AATAAATACA TACATACATA CATACATACA TACATACATA CATACATACA TACATATGAG 20400
GTTGACCGAT TGCTAATCCT TTAGAATTTT GTTTTAAAT AATGTGATAG AGAGATAGAT 20460
AATAGATAGA TGGATAGAGT GATACAAATA TAGGTTTTTT TTTAGTAAA TATGAGGTGT 20520
AATTAACACT TTTCCTTTTT TAGGTTTTTT TTTTTCCTCC CTGTCCATGT GTTGTGCTGG 20580
AITTGAACCT AGGACCTTGG CAGGTCAACT GGAAAACGTG TTTTCTATAT ATATAAATAG 20640
TGCTCTGTCT GCTGTTTGTG TGTGTTGTGT CTGTGCTGTG TGCTTGCTGT CTGTGCTGTCT 20700
TGCTTTTTTT TTCTCTCGA GACAGTATTT CTTCTGTATA CCTGGTGCCC TGAAACTTAC 20760
TCTGTAGACC AGCCTGGCCT CAATCGAACT CAGAAATCCT CCTGCCTCTT GTCTACCTCC 20820
CAATTTTGGG GTAAAGGTGT GCTACACCAC TGCTGCGCAT TATTATCAT TATTATTTA 20880
ATTTTATAT TAGACAGAAC GAAATCAACT AGTTGTCTCT GTTTCGTATA TTTATTTGAA 20940
AATTAGTTGA CCAATTAGTT GCGTGGTTTG GGAGGTTTTT TTTGTTCCG ACTTTGGGTG 21000
TTGTGGGGCT GGGGATCAGG TATCTCAACG GAATGCATGA AGGTTAAGGT GAGATGGCTC 21060
GATTTTTGTA AAGATTACTT TTCTTAGTCT GAGGAAAAAA TAAAATAATA TTGGGCTACG 21120
TTTCATTGCT TCAITTTCTT TTCTTTTCT TTCTTTCTTT CTTCAGATA CTTCAGATA 21180
CCAGTCTCTG CGACTTCTG GAAGATGTAG GCAITTCAT TGGGAAAAGCA TTGTTTGAGA 21240
GATGTGCTAG TGAACCGAG AGTTTGGATG TCAAGCCGTA TAATGTTTAT TACATATAG 21300
AAAGTCTCTA CAAAGGTGT CTTTAACTTT TTTTTTTTT TTTCCTCTC TACTCTACT 21360
TGTCTCCTAC CTCCACCAA CGCGCTTGT ACATGATG TGAGCTTTGT TTTGCTTAC 21420
AGACATATAT TTTTCTTTT GGTTTGCTT GACATGTTGT CCTCTTCTAT CCGTCTTAC 21480
TTCCAGACG GCTCTTTGAG AATAAATGG GAGCGCAGAA CCAAGCTTTT TTGAATAAG 21540
CACCACAAC CTAACTCTTT TGGCTGTTTT CCTTCCCAAG GCACAGATCT TTCCACGAT 21600
GGAAAACAT GTAGCAGTTG TAGGACACAC TAGACAGAG CACCAATCTC CATTGTGGT 21660
GGTGTGTAAC CACCACCAT GTATTTGGCT GGGATTGAA CTCAGATCTC TCAGAAGAC 21720
AGTCAACCAT TAAACCGAT GAGCCATCTC TCCAGCCCTC CACCATCTCT TCTTAAGCA 21780
TGAATGATCC CAGCATGGGA AGACAGTCTG CCTCTTTTGT GGTATATCAC CATATACTA 21840
ATAAATATTC GAAATGAATG AAGTCTCCAC GTATTATTTT CTTCGAGTCA TCTAAATCT 21900
CTCCAGACAC CTCCCCCTCC CCCCACTGCT CTTTCTCCCT ATGTTTGGGT GGGGCTGGG 21960
GAGGGTGGG GTGGGGGCGG GGATCTGCAT GTCTTCTTGC AGGCTGTGGA ACTATTTGCG 22020
ATGGCTGGT TCTCTGAAT GTTGGACCTT GTCTATCCAG AGGCTGACTG GCTAGTTTTT 22080
TACCTGAAGT CCGTGAATGA TGATTTCCCT GTGAATTC 22118

```

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42999 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCTGACACGC	TGTCCTCTGG	CGACCTGTGC	TCCGAGAGGT	TGGGCTCTCG	GATGCGCGCG	60
GGGCTCTGGC	CTACCGGTGA	CCGCGTAGCC	GGCCCGGCTC	GTCGCTTGAG	CCGCTCTGCC	120
CGCCCGCGGG	GCGCTGCTGT	CTCTCGCGCG	TCCGACGGTC	CCGACTCCCG	GTGCGCGGCC	180
GGTCCCGGGT	CTCTGACCCA	CCCGGGGGCG	GGCGGGAAGG	CGCGGAGGGC	CACCGTCCCC	240
CGTGCCTCTC	CGCGTGCGGG	CGCCCGGGGC	GCGGCACAAAC	CCCAACCCGT	GGCTCCGTGC	300
CGTGCCTCTC	AGGCGTTCTC	GTCTCGCGGG	GGTTGTCTCGC	CGCCCTTTCC	CGGAGTGGG	360
GGGTGGCGCG	AGCCGATCGG	CTCGCTGACC	GGCCGGCCCTC	CGCTCCCGGG	GGGCTCTTCG	420
ATCGATTGGT	TGACGTCTGT	CTCTCCCGGG	CGGGGTCCGA	GCGCGACGGC	CGAGAGGGCG	480
GCAGTTTCGT	GCGAACGGGA	CCGTCTCTCT	CGCTCCCGCC	GCGCGGTCCC	CTCGTCTGCT	540
CTCTCCCGCG	CCCGCGGGCC	GGCGTGTGGG	AAGGCGTGGG	GTGCGGACCC	CGGCGCGACC	600
TGCGCGTCCC	CGCCGCGGCG	TTCGCTTCGC	GGGTGCGGGC	CGCGGGGGTC	CTCTGACGGC	660
CGCAGACGAC	TGCGCTGTGC	CTCCAGTGG	TGTGCGACTT	GTGCGCGGCC	CCCTTCGCGG	720
GGGTTGGGGG	GCCCGTCTGC	CGCGCCGGT	GTGCTGACCT	CTCGGGGGGG	GTTTGGCGGA	780
GGTTCGCGCT	GCGCTGCGTC	CTTCGCGTCC	TCTCGGAGCG	CTCCGGGTTC	TCCCTCAGGT	840
GCGCAGAGGC	GAAAGGTGCT	GTGTCTCTCC	CGCCCGGGCG	GCCCCCTCT	CGGTTCGCGG	900
CGCGGTGTTC	CGCGCTGGG	TCCTGAGGGA	GCTCTCGGCT	GTGGGGTTTC	AGCGCGTTTC	960
AGTGAGAGCA	GACGAGACGC	GCCCCCTCCA	CGCGGGGAAG	GGCGCCCGCC	TGCTCTCGGT	1020
GAGCGCACGT	CGGTGTCTCC	CCTCTGGCGG	GTGCGCGGGG	GCGGTGTGAG	CGATCGCGGT	1080
GGGTTCCGGG	CCGCTGTACC	CGTGGCGCGG	CGCGCCGCGC	AGGGGCTGCC	GTTTCTGCGCT	1140
CGACCGGTGC	TGTTGTGGTT	GACTTCGGAG	CGCTCTGACC	TCCGAAGGAA	GCGAGTGGGT	1200
GGACCGGGTC	GCGTGTGGTG	GTTGCGCGCA	CGCGCGCACG	CGCCGGGCGC	CGCGCTTGA	1260
CGCGAAGCCT	CGAGGTGGCC	GCGCGCAGGT	GTTTCTCTGT	ACCGCAGGGC	CCCCCTCCCT	1320
CCCCAGGCGT	CCCTCGGGCG	CTCTGCGGGC	CGAGGAGGGA	GCGGCTTGGC	GTGGGGGGGA	1380
GTGTGACCCA	CCCTCGGTGA	GAAAGACCTT	CTTAGAGCAT	CTGAGAGGCG	TGCCCTTGGG	1440
GTACCGGATC	CCCGGGGCGC	CGCCCTCTGT	CTCTGCTTCC	GTTATGTAGT	CGCTGCCGTA	1500
GGACCCGCGT	CGAGAGGAGC	CCTCTCCCGC	TTCGCCCTCG	ACGGGGTTTG	GGGGGAGAAG	1560
CGAGGGTTTC	GCGCGCCACC	CGGTTGGTGG	CTCTGCGGCT	ACTGTGGACC	ACTGTGGACC	1620
CGCGCTCCCC	CTTCCGAGTC	GGGGGAGGAT	CCCGCCCGGC	CGGGCCCGGC	GCTCCCAACC	1680
AGCGGGTTTG	GACGCGGGCG	CGGCGGGGCG	GTGGGTGTGC	GCGCCCGGCG	CTCTGTCCGG	1740
CGCGTGACCC	CTCTCGTCCG	CGAGTCGGCT	CTCCGCGCGC	TCCCGTCCCG	AGTCGTGACC	1800
GGTGCCGACG	ACCGCGTTTG	CGTGGCACGG	GGTCGGGGCC	GCTTGGCCCT	GGGAAAGGAT	1860
CCACCGGTGG	GGGCGCGCCG	GTCCTCCGGA	GCGGGACCGT	GTGCGAGGAT	GAGCAGGAAT	1920
CACGAGCGAC	GGTGTGGTGG	CGGTGTCCGG	TTCGTGGCTG	CGGTTCGCTC	GGGGCCCCCG	1980
GTGGCGGGCG	CCCGGGGCTC	GCGAGGCGGT	TCTCGTGGGG	GCGCGAGGGC	CGTCGCGCGT	2040
CCCAGGCGGG	GCGCGCGGGG	ACCGCCCTCG	TGCTGTGGGG	GGTGGGATCC	CGCGCGCGTG	2100
TTTTCTGGGT	GGCCCGGGCG	TGCTGAGGTT	TTCCTCCCGA	GCGCGCGGCT	CTCGGGGACT	2160
CGGGTGTGCC	TGCGCTTCGC	GGTCCCCGCG	CCTCGCCGCT	CTGTGCCCTG	TTCGCGGCC	2220
CGCGCCGCGC	GATCCTCTTC	TTCGCCCGGA	GCGGCTCACG	GGCTTCACGT	CGGTGTGGTG	2280
CCCCGCTTGG	GCGCAGACCC	CGTGGGGCGC	CGTGGGGCGC	CGCCCGCGCG	CACCTGATCG	2340
CCCGCGCTTC	CGGTCCCGCG	GCGCGCGGCT	TGGGAGACGG	GTGCGTGGCG	CGCGCGGTGC	2400
GGCCCGGTGG	GCTTCCCGGA	GGTTCGCGG	GCTCGCGCTG	GCGCGCTGCG	GGGGGAGGAG	2460
ACGGTTCGCG	GCGACCGGCT	CGGCTTCGCG	CGCGCGCGGT	CGCGCGGCGG	CGCGTGAAGC	2520
TCCGCGAGGG	CGCGTGGGCC	GCCCCGGGTC	CCCCCGCGTG	CGCGCGCGCG	CGGTGAGGCC	2580
CCCGCGGTGT	GTCCCGGCTG	CGGTGCGCGG	CGTCCGAGGG	CTCCCGCTGG	CGTCCCTTTC	2640
CCCGCGGAGC	CGCTTTCTCG	CGCCTTCCCG	GTGCGCCCGG	CCTCGCGCGT	GGTCTCTCGT	2700
CTCTCTCCGG	CGGCTCTCTG	GAAACCGGGT	CGCGCGTCCG	CCCGGGTGGC	CTCGCTTCC	2760
CGGGCTCACC	CGCGCGCTTC	CCCGAGGCGT	CGGTCCCGGG	CGTCCGCGCT	GGGAGAGGCT	2820
CGTCTCGCGC	GCGTGGCGTC	GCCCCTTTCG	GCGCGCGGCT	GCGCCCGAGC	CGCGCGCGCG	2880
GGTCTCCTCC	GGAACAGCGT	TGCTGCGACG	TGTGGGCTGG	GTGACCTCTC	GCTTTCGCGC	2940
TGCTCTGCCC	CTTCCCGGCT	TGGGGGGGTG	GGGCCCGGGC	CGGGGCTTCG	GCCCCGGTGC	3000
CTGCTCTCCG	TCCCGGGGCG	GGGCGGGCGC	GCCGGCCGCG	CTCGGTGCGC	CTCTCTTGGC	3060
CGTCTGTGGT	CGTGTGGCAC	CCCTGCGCGG	GCGCCCGCGC	CGCGGCTTCG	GAGCCGGGCT	3120
TGCGCGCGGG	CCCGGCGCCT	GAAACGGAAC	GGCTGCGCGG	CGGCTGCGCG	CGCACGGGCG	3180
GACTGTCTCC	GCGCGCGGGA	CCGCGGTTCG	CTCTCTGCTC	GCGCCCGCGA	CGCGGGGCGC	3240
GCCCCCGGGG	CGCGGCGGAG	GCGCGTCCCC	GCTTCCGCGC	GCGCCCGGGG	GCGCGCGCGC	3300
GCGCGCGCGC	CGGTGGCGCG	CGGTCTCTCC	GCGCGCGCGG	CGCGGGTTCG	CGGCTCTCCG	3360
CTCTCTCGCG	GCGGGCGCGA	GAAAGAACGG	TGCGGGGTCT	GTGGCGCGGG	GCCCCCGGGT	3420

GTGCTGTGCG GTGGGGGGGG GGTGGTTGGG GCTGCGGTT CGCGCGGCC CGCCCCGGCC 3480
 CCACCGGTCC CGGCGCGCGG CCGCGGCCGC GCTCGCTCCC TCCGTCGCG CGGTGCGCGG 3540
 CCGCTGCGTC CGTCCGTGCG TCGTCTCTCT CGCTTGCGGG CGCGCGGGCC CGTCTCTCGG 3600
 AGCGCCCGCG CGCGGCCGTC CGCGCGCGTC GGGGGCTCG CGCGCTCTAC CTTACTTACC 3660
 TGGTTGATCT TGCCAGTAGC ATATGCTTGT CTCAAAGATT AAGCCATTGA TGTCTAAGTA 3720
 CGCAGCGCGG GTACAGTGAA ACTGCGAATG GCTCATTAAG TCAAGTTATG TCTCTTTGGT 3780
 CGCTCGCTCT TCTCTACTTT GGATAACTGT GGTATTTCTA GAGCTAATAC ATCCGCGACG 3840
 CGCGTGAACC CCTTCGCGGG GGGGATGCGT GCATTTATCA GATCAAAACC AACCGGGTGA 3900
 GCCCTCTCTC CGCCCCGGCC GGGGGCGGG CGCGCGGGCC TTTGTTGACT CTAGATATCC 3960
 TCGGGCCGAT CGCACCGCTC CCGTGGCGGC GACGACCCAT TCGAACGTCT GCCCTATCAC 4020
 CTTTCCGATG TAGTTCGCCCT GCCTTACCATG GTGACCCAGG GTGACCGGGA ATCAGGGTTC 4080
 GATTCCGATG AGGGAGCCCTG AGAAACGGCT ACCACATCCA AGSAGGCGAG GACGGCGCGA 4140
 AATATCCCAT TCCCGACGCG GGGAGTAGTT GACGAAAATT AACATACAG GACTCTTTCG 4200
 AGGCTCTGTA TTTGGAAATG GTCCACTTTA ATCTCTTTAA CGAGGATCCA TTGAGGATCA 4260
 AGTCTGTGTC CGACGACCGC GTTAAATCCA GCTTCAATTA AGTTCCTGCA 4320
 GTTAAAAAGC TCGTAGTTTG ATCTTGGGAG CGCGCGGGCG GTCCCGCGCG AGCGGAGCCA 4380
 CGCCCGCTCC CGCCCGCTTG CCTCTCGGG CCCCCTCGAT GCTCTTAGCT GAGTGTCCCG 4440
 CGGGCGCCGA ACGCTTTACT TTGAAAAAAT TAGAGTGTTC AAAGCAGGCC GTAGCCGCTC 4500
 GGAATCCGCA CTTAGGAATA ATGGAATAGG ACCCGGGTTC TATTTTGTGT GTTCTCGGAA 4560
 CTGAGGCCAT GATTAAGAGG GACGCGCCGG GGCATTCTGA TTGCGCGCGT AGAGGTTGAA 4620
 TTCTTTGACC GCGCAACAGC GGACCAAGAG GAAAGCATTG GCCAAGATTT TTTTCAATTA 4680
 TCAAGAAACGA AAGTCGAGG TTCGAAGACG ATCAGATACC GTCTAGTTCT CGACCATATA 4740
 CGATCCCGAG CGGAGCTGCG CGGCGTTTAT TCCCATGACC CGCGCGGGCAG TCTCCGGGAA 4800
 ACCAAAGTCT TTGGGTTCCG GGGGGAGTAG GTTGTCAAAG CTGAAACCTA AAGAAATTGA 4860
 CGGAAGGCGA CCGACAGGAG TGGAGCTGCG GGCCTTAATT GACTCAACAG GGGAACTCTG 4920
 ACCCGCCCGC GACACGGACA GGATGACAG ATTGATAGCT CTTTCTCGAT TCCGTGGGTT 4980
 GTGGTGCATG GCTGCTCTTA GTTGGTGGAG CGATTTTGCT GGTTAATTCC GAGCAAGAAC 5040
 GAGACTCTGG CATGCTAAT AGTTACGCGA CCCCAGAGCG GTCCGCGTCC CCAACTTTCT 5100
 TAGAGGGACA AATGGGCTTC AGCCACCCGA GATTGAGCAA TAACAGGTCT GTGAGTCCCT 5160
 TAGATGTCCG GGGCTGACCG CGCGCTACAC TGACTGGCTC AGCGTGTGCC TACCTTACCG 5220
 CGCGAGCGCG GGGTAAACCG TTGAACCCCA TTCGTGATGG GGATCGGGGA TTGCAATTAT 5280
 TCCCATGAAA CGAGGCAATT CCCGAGTAAG TCGGGGTGAT AAGCTTTGCT GATTAAGTCT 5340
 CCTGCGCTTT GTACACACCG CCGCTCGCTA CTACCGATTG GATGGTTTAT TGAGGCCCTC 5400
 GGATCGGCCC CCGCGGGGTC GGCCACCGCG CTTGCGGAG CGCTGAGAAG ACGGTGCAAC 5460
 TTGACTATCT AAGGGAAGTA AAGATCGTAA CAAGGTTTCC GTAGGTGAAC TCGCGGAAGG 5520
 ATCATTAACG GAGCCCGGAG GGGGAGGCC CGCGCGGGCG CGCCCGCGCC CGCGGCTTCC 5580
 TCCCGCACAC ACCCGCCCCC CGCGCGACCG GCGCGCTGCG CGCGCGGGCG CGCGGCTGCC 5640
 GTTCGTTCCG TCGCTCGTTC GTTCGCGGCC CGCGCCCGCC CGCGCGAGAG CGGAGAACTC 5700
 GGGAGGGAGA GGGGGGGGAG AGAGAAGAG AGAGAAGAG AGAGAAGAG AGAGAAGAG 5760
 AGAAGGGCGT GTCGTTGGTG TCGCGTGTCT GTGGGCGCG CGCGCGCGCG GAGCGCGTCC 5820
 CGCGCCGCGC CCGCGACGAG GTGGGTGTG CGCGCGCGCG GCGCGGTCTT CGCGCGGTCT 5880
 CGCGCGGTTC CGCGCGGTTC TCGGTGCTCT CCGCGCGCGT CGCGCGGTCT GTTCGCGGCC 5940
 CGCGCGCGCG CGCGCGCGCG CCGCTCGCTC TCCCGGGCTT TCCCGTCCG CGCGCTCGAG 6000
 GTTCGCGGCG CGGACCGCGG TCCCTCCCC CGCTCTCTCG TCCCGCCCC CGCGCTCCAG 6060
 GTACTTAGCG CFTTCCGCG CGGAGGTTTA AAGACCCCTT GGGGGGATCG CCGCTCGCG 6120
 CGTGGTGGCG GGGCGGTGTG GGGCCCGCGG GGGAGTCCCG TCGGGAGGGG CCGCGCCCTC 6180
 CCGCGGCTTC CACCGCGGAG TCGCTCCCC GCGCGGGGCG CGCGCGCGCG CCGCGCGCGG 6240
 CGCGCGCTCG GGTGGGGGCT TTACCCGGCG CGCTCGTGG CGCTGCGCGT CGTGTGCGGT 6300
 CGCGCCCGCG CCGTGGGGGCG GGGAAACCCC GGGCGCTCGT GGGCGGTGTG CGCGCTTCG 6360
 CCGCGCGTGG CGCGCGCGCG CCTCCCCGTG GTGTGAACC TTTCCGACCC TCTCCGGAGT 6420
 CGGTCCCGGT TTGCTGTCTG GTCTGGCGCG CTTAGAGCAA CCGCTCTCTT CTGTTGGCGG 6480
 GGGGGCGGGG GGGAGCTGCC GCGCCAGGAA GGGGCTCTCT CGCTGCGTCT GTGCGGAGCG 6540
 CCTCGCCAAA ATCGACTCTT TACGACTCTT AGCGGTGGAT CACTCGGCT GTGCGTGTAG 6600
 GAAGAAGCCA GCTAGCTGCC AGAATTAATG TGAATTCGAC GACACATTGA TCATCGACAC 6660
 TTGAAACGCA CTTGCGCGCG CGGGTTCCTC CCGGGGCTAC GCTGTCTG GCTGCGCTTG 6720
 CGGATCAATC GCCCCGGGGG TGCCCTCCGG CTCTCGGGG TCGCGGCTCG GGGGTCTCCT 6780
 CGCAGGGCCC CGCGGGGGCG CTCGCTCCCC CTTAAGCGAG ACCCGCGCGG TCTGCGCCCT 6840
 CTTTGGCGCG CGCGCGCGCG CCTTCCCCCT CCCCCCGCG GCGCTGCGTG GTACCGCGCT 6900
 GGGTGGCGGG GGGGAGAGGG GGGCGCGCCC GGGCTGAGAG GAGGAGGAG GCGCGCGCG 6960
 CGCCGGAAGA CGGAGAGGGA AAGAGAGAGC CGGCTCGGGC CGAGTTCGCG GGGCCGCGG 7020
 TCGCGTCCG GTTTCCTCC TCGGGGGGCG CCTCGCGCC CCGTCCGCTG GTGCGTCTG 7080
 GGTTCGTGCG CCGCGCGCGG GTGGAAGGTT CGGTGCGGCT TGGTGTGGG GGAGGAGGAA 7140
 GTCCGCGGTG GCGCGGTGTT GGTGCGGTG TGGTGTGGG GGAGGAGGAA GTGCGGTTCG 7200
 GAAGGGGAAG GGTGCCGGCG GGGAGAGAG GTGCGGGGAG GCGGTCCCGG TCGCGCGGCT 7260

TCCGCGCGCC	GCCCCGGGT	GCGGCGCGG	GTCCGCGCGA	C CGGCGCGCT	C CGGCGCGCC	7380
TCCTCTCTCC	CGCGCGCCCT	CCTCCGAGCG	CCGCGCGGTC	CTCTTCGCGC	TCCCGCGCGG	7440
TACGGCGCGG	CGCCCGCGCG	CCCGGCTCGC	CTCGCGGCGC	GTCTCGCGCG	GCGGGGAGCG	7500
CGCCCGCGCG	CGCGCGCGGT	CGCGCGGCGC	CGGGGTTCGC	GTCTCCCGCG	GCGCGACCGC	7560
CGGAGCGCGG	CGGTGTGCTC	CGCGCTCGCG	CGCCCGGCTC	CGGCTCGCGG	CGCGCGCGCG	7620
CGCGCGCGGG	GCCCCGTCCC	GAGCTTCGCG	CTCGGGGCGG	CGCGGCTCGC	CGCGCGCGCT	7680
CTCGGACCGG	TCCCGCGGAC	CTCCGCGGGG	GAGACGCGCG	GGGGCGTTCG	CGCGCGGCTC	7740
CGCCTCCGCG	CGCTGCGCTC	CCCTCCGCTC	GTCCCGCTCC	GGCGGGGCGG	CGCGGGGCGG	7800
CGGTGCGGCG	CGCGCTCTCT	CTCCCGTTCG	CTCTCCGCTC	CGCCGGGCGC	GTCTCCCGAC	7860
GGAGCGCTCG	GCGGGCGGCT	GCGCGGCGCG	GATTCTCGTC	GTCTCTCGCG	CAGAGCGGCC	7920
GCTCCCGTCC	GAGACGCGAC	CTCAGATCAG	ACGTGGCGAC	CGCTGGAATT	TAAGCATATT	7980
AGTCAAGTGA	GGAAAGGAAA	TAACACGAGA	TCTCTCAGT	GAAACGAGAT	TAACAGGGAA	8040
GAGCCAGCGG	CCGAATCCCC	GCCCCGCGGG	GCGCGGGACA	TGTGCGGTAC	GGAAGACCGG	8100
CTCCCGCGCG	CGCTCTGTCG	GGGGCCCAAG	TCCTTCTGAT	CGAGGCCGAC	CCCGTGGACG	8160
GTGTAGCGCG	GGTAGCGGCG	GGCGCGCGCC	CGGGTCTTCC	CGAGTTCGGG	TTGCTTGGGA	8220
ATCGAGCCCA	AAGCGGGTGG	TAAACTCCAT	CTAAGGCTAA	ATACCGGCAC	GAGACGGGATA	8280
GCTCAACAAT	ACCGTAAAGG	AAAGTTGAAA	AGAACTTTGA	AGAGAGAGTT	CAAGAGCGCG	8340
TGAACCGGTT	AAGAGGTAAA	CGGTTGGGTT	CGCGCGAGTC	CGCCCGAGGT	ATTCAACCGC	8400
GCGGGCGGCT	CGCGCGGTTC	GGCGGCGCGG	CGGATCTTTC	CGCCCGCCCG	TTCTCTCCGA	8460
CCCTCTCACC	CGCTCTCCCT	TCCCGCGCGG	CCCTCTCTCC	TCCTCCCGCG	AGGGGGGCGG	8520
CTCCCGCGGG	TGCGGGGGTG	GGCGGGGCGG	CGCGGGGGTG	GGGTGCGGCG	GGGACGCTCC	8580
CCCGACCGCG	GACCGCGCGG	CGCGGGGCGG	ATTTCACGCG	CGCGGTTGCG	CGCGGACCGC	8640
CTCCGGGACG	GCTGGGAAGG	CCCGCGGGGG	GAAGTGGCTC	GGGGGGCGCG	GTTCGTCGCT	8700
CGCTCTCTCT	CGCTCGCGCT	CGCGCGCGCG	CGCGCGCGCG	TCCTCTCTCG	GCGGCGGCGG	8760
CGCGCGCGCG	CGCGCGCGCG	CGCGCGCGCG	CGCGCGCGCG	CGCGCGCGCG	GCGACCGGAA	8820
CGCCCGCGGA	GTGTTACAGC	CGCCCGCGCA	CGACGACTCG	CGGAATCCCG	GGGCGGAGGG	8880
AGCGAGACCC	GCGCGCGCGC	TCTCCCCCTC	CGCGGCGGCC	ACCCCGCGCG	GGAAATCCCG	8940
GCGAGCGGGG	TCTCCCCCGC	GGGGGCGCGC	CGCGGCTCTC	TGCTGGGGGG	CGCGGGCCAC	9000
CCCTCCACAG	CGCGACCGCG	TCTCCACACC	CTCTCCCGCG	CGCCCGCCCG	CGCGGACGCG	9060
GCGGGGCTCC	CGCGCGCGGT	CGGGGGGCGG	GGCGGACTCG	CCCCAGTTCG	CCCGGCGCGG	9120
CTCGCGCGGT	CGGCGCGGGG	GGAGGTTTCT	TGCGGGCCAC	CGCGCGCTCG	CCCGAAGAGG	9180
GGGAGCGGGG	ACCGAGCGCA	CGGGGTTCGC	GGCGACGTCG	GCTACCCACC	GACCCGCTCT	9240
TGAACACGGG	ACCAGAGGAT	CTAACACGTC	CGCGAGTTCG	GGGCTCCGAC	GAAAGCGCGC	9300
GTGGCGCAAT	GGAAGTGAAG	GCGGGCGCGC	TGCGCGGCGC	AGGTGGGATG	CGAGGCGCTC	9360
TCCAGTCTCG	CGAGGGGCGA	CCACCGGCCC	GTCTCGCCCG	CGCGCGCGGG	GAGGTGGAGC	9420
ACGAGCGCAC	GTGTTAGGAC	CCGAAGAATG	GTGAACATAT	CTTGGGCAAG	GCGAAGCCAG	9480
AGGAACATCT	GGTGGAGGTC	CTGAGCGGTC	CTGACGTGCA	AATCGGTCGT	CCGACCTGGG	9540
TATAGGGGCG	GAAGACTAAT	CGAACCATCT	AGTAGCTGGT	TCCCTCCGAA	GTTTCCCTCA	9600
GGATAGCTGG	CGCTCTGCGA	GACCCGACGC	ACCCCGCGCA	CGCAGTTTTA	TCGGTAAAG	9660
CGAATGATTA	GAGGTCTTTG	GGCCGAAGCG	ATCTCAACCT	ATTCTCAAAC	TTTAAATGGG	9720
TAAGAAGCCC	GGCTCGCTGG	CGTGGAGCGG	GGCGTGAAGT	GCGAGTGCCT	AGTGGGCCAC	9780
TTTTGGTAGG	CAGAACCTGG	GCTCGGGATG	GAACCGAAGC	CGGGTTAAG	GCGCGGCTAG	9840
CGGACGCTCA	TCAGACCCCA	GAAAGGTTGT	TGGTTGATAT	AGACAGCGAT	AGCGTGGCCA	9900
TGGAGATCGG	AATCCGCTGA	GAGGTGTGTA	ACAATCTACC	TGCGGAATCA	ACTAGCCCTG	9960
AAAATGGATG	GCGCTGTAGC	GTCGGGCCCA	TACCGGGCGC	TGCGCGGCGG	TCGAGAGTGG	10020
ACGCGAGCGG	CGGGGGGCGG	GCGCGCGCGC	GCGCGTGTGG	TGTGCTGTCG	AGGGCGGCGG	10080
GCGCGGCGCG	GCGCGGCGGT	TGGGGTCTTT	CCCCCGCCCC	CGCCCGACCG	CTCTCTCCCC	10140
TCTCTCCGCG	CACCGCCCCG	TCCCGCGCCC	CGGAGCGCGG	CGGAGCGCTAC	CGCCGCGACG	10200
GTAGGAGGCG	GCGCTCGGTT	AGCCTTGAAG	CCTAGGGGCG	GGGCGCGGCT	GAGCGCGCGC	10260
CGGTCTGAGA	TCTTGTGTGT	AGTAGCAATG	ATTCAAGGCA	GAACCTTTGA	GCGCGAGGCT	10320
CAGAGCGGTT	ACGAGTGTGA	CATGGTCTAG	TCGGTCTGCG	TCGGTCTGCG	GAGATCTGCG	10380
AGCGCGGTTT	CGAGGCGAGC	GCGGATGCGC	TCGCTTGCCG	TCGCGCGGAT	GAAAGGAGAT	10440
CGGGTCTCAG	TCCCGGAATC	CGGAGTGGCG	GAGATGGGCG	CGCGGAGCGG	TCCAGTCCGG	10500
TAACGCGACG	GATCCCCGAG	AAGCCGCGCG	GAGCGCCGCG	GAGAGTTCTC	TTTTCTTTGT	10560
GAAAGCGTAG	CGCCCTTGGA	ATGGGTTTCG	CCCGAGAGAG	GGGCCCCGTC	CTTGGAAAGC	10620
GTGCGGTTTG	CGCGGGGCTC	CGGTGAGCTC	TGCGTGCCTC	TGTAATAATC	GGGGGAGAGG	10680
GTGTAAATCT	CGCGCGCGCG	CGTACCATAA	TCCGACGACG	GTCTCCAAAG	TGAACACGCT	10740
CTGGCATGTT	GGAACAATGT	AGGTAAAGGA	AGTCGCGAAG	CGGATCCGCT	AACCTTCGGA	10800
TAAAGATTTG	CTCTAAGGGG	TGGGTGCGTC	GCGCTGGGCG	GCGAAGCGGG	GCTGGGCGCG	10860
CGCGCGGCTG	GAGCGAGGCG	GCGCGCCCCC	CCACGCCCCG	CGGACCCCGC	TGCGCGGCTC	10920
CCCCCGCCCC	ACCCGCGCGC	GCGCGTCGCT	CCCTCCCCAC	CCCGCGGCTC	CTCTCTCTCT	10980
CTCTCCCCCG	TCTCCCCGTC	TCCCCCCCTC	CGGGGGGAGC	GCGCGTGGGG	GCGCGGCGCG	11040
GGGGGGAAGG	GTCGGGGCGG	CAGGGGCGCG	GCGGCGGCGC	CGGGGGCGCG	GCGCGGGGCG	11100
AGGTCCCCCG	GAGGGGGGCG	CGGGGACCCG	GGGGGGCGCG	CGGCTTGAGC	GACTCTTGAG	11160
GCGAGCGCGG	CCCTTCCCGT	GGAATCGCCC	AGCTGCGCGG	GCGCTGCGCG	CGCCCGCCCG	11220

-199-

CAGCTGCCCG CGCGGAGCTC CGGACTTAGC CGGCGTCTGC ACGTGTCCCG GGTCCAGCAG 15180
 CAGCGCGCCG CGGACGCTGC CGGCGCACCG ACGCGAGGGC GTTCACGCGC GTTTCACGCGC 15240
 CGGCGACCTC CACCGGCTCT CGCCCGCGGT GGAGCTGGGA CCACGCGGAA CTCCTCTCTC 15300
 CACATTTTTT TCAGCCCCAC CGCGAGTTTG CTCCGCGGG ACTTTTAAGA GGGAGTCACT 15360
 GCTGCCGTCA GCCAGTAATG CTCTCTCCTT TTTTGTCTTT TGGTCTTGCC TTGCGTTTTC 15420
 TTTCTTTCTT TCTTTCTTTC TTCTCTCTCT TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC 15480
 TCTCTCTCTC TCTCTCTCTG TCTCTCTCCC CTCCCTCCCT CCTTGGTGCC TTCTCGGCTC 15540
 GCTGCTGCTG TCGCTCTTGC CTCACGGTT CAAGCAAACA GCAAGTTTTC TATTTTCGAGT 15600
 AAAGACGTAA TTTCACCAAT TTGGCCGGGC TGGTCTCGAA CTCGCCACCT AGTGAATCCG 15660
 CCGCTCTGCG CCGCCCAAAGA CTGCTGGGAG TACAGATGTG AGCCACCATG CCGCGCCGAT 15720
 TCCTTCTCTT TTCAATCTTT ATTTTCTGAA CGCTGCCGTG TATGAACATA CATCTACACA 15780
 CACACACACA CACACACACA CACACACACA CACACACCCC TATGATGATA 15840
 AACTATGTAA ATGATATTTC CATAATTAAAT ACGTTTATAT TATGTACTTT TTAATGTGATG 15900
 AATATGATTC GAAGCCCAAT TTCAATTACA TACACGTGTA TGTTATCTCT TCTCCCTCTC 15960
 CTTTCAATCAT TATTTATTAA TAATTTTCGT TATTTTATTT TCTTTTCTTT TGGGCGCGGC 16020
 CCGCTGTGTC GGGCTGTGCT TAAGCCCGGG AGGATCTCTC TCCCAATAG CTGGGACTAC 16080
 AGGATCTCTC GGGCTGTGCT TCGCTCTCAAT TGCTTTTATG ACCTGTGAT TTGCTGTGTT 16140
 CAGCTTACGT GGGCTGTGCT TCGCTCTCAAT TGCTTTTATG ACCTGTGAT TTGCTGTGTT 16200
 CGATTGATTG CGATCTCAAT TGCTTTTATG ACCTGTGAT TTGCTGTGTT 16260
 ATTCTCATGG GTCTCTCTGT GTCAITGTICA GTTCAATCTG TCGTCTGCTT TCGTCTGCTT 16320
 TTAITTCCTGT CCTCTCTCTG TCTCTCTCTG TCTCTCTCTG TCTCTCTCTG TCTCTCTCTG 16380
 CTGGAGCTGT CTCTCTCTGT CTCTCTCTGT CTCTCTCTGT CTCTCTCTGT CTCTCTCTGT 16440
 CGACCAAACG CTCTCTCTGT CTCTCTCTGT CTCTCTCTGT CTCTCTCTGT CTCTCTCTGT 16500
 CGACCAACCA CACCGGCTGA CTTTATGTGT GTTCTCATGT TTTTCCGTAG GTAGCTATGT 16560
 GGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT 16620
 ATGATGTATC GTATGTATGT ATGTATGTGA GTAGATGCG TTTCCGGGTT CTATCATGTT 16680
 CGGCACTGCT CTTCTGAATC CCGTCTCTCA AGCAATCCGG CTGCTGCTT GCGCGGCCCA 16740
 CACTGTCTCT ATTACAGCGC TGAGACGCTG CGCCTGGCTC CTCTACATCT TGCTGTGCTT 16800
 CCTGCTGCTT TCGGCGCTA TCAATCGTCT TCTTTTATG ACGGATGTCT TGCTGCTTGA 16860
 TTGTTCATCT TCTGGGCAAC CGTGGTCTCT TTTCAAATCT CTATGATTAT TATTATTGTA 16920
 GGGCTCATGC CACGTGTGCA GGTGATCTCG AACTTTTAGT CTCAGAGATC CTTCCGCACT 16980
 CGGCTCTCGG GAGTGTCTGT ATGACACGCG TGGGACGCTG ACGCTCTGTT CGTGTTTGTC 17040
 GGGGTTCGTT TCTTTTCGTT TTTAATACGG GGACTGCGAA CGAAGAAAT TTTACAGACG 17100
 ATCTCACGGA TCGGCTTTT GTTCTTTTCT TTTTATTTCT TTTAGACGGA GTTCTCACTT 17160
 GTGCGCCAG TGGTGAGTAC GATGGCGGCT CGCGGCTCAC GCGACCTCTC GCTCCCGAGT 17220
 TTCAGATGAT TCTCTGCTCT CAGCCTCTCC GAGTAGCTGG AATGACAGAG ATGAGCAGTC 17280
 GTGCGCGGCT AATTTTCTTA TTTTGTATG AGATGGGTTT TCTCCATCTT GGTCAAGCTG 17340
 GTCTTCAACT TCCGACCGTT GAGAAATCTT AACTTTCTGT GTGGTGTGTT TTTTCTTTT 17400
 TCTTTTCTCT TCTTTCTCTT TCTTTCTCTT TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC 17460
 CTTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC 17520
 CTTTCAAGCT GGCCTCTCTA CTTGTGTTGC TCTGTGTTGC TCTGTGTTGC TCTGTGTTGC 17580
 CGCTTGACTC TTTCCCGCTC ACATCCGCGG TCTGGTTGTT GAAATGAGCA TCTCTCGTAA 17640
 AATGGAAAG ATGGAAGAAA TAAACACGAA GACGGAAGAG ACGGTGTGAA GTTCTCTCTT 17700
 GCGCTCTCCC GGGGGTGATC TTGGACCGG AAACACGGAG GAGGCTTGGC TGAGTGGGTT 17760
 TCTGGTGCCG AAGGCTCCCG AGGCGCTCTC TCTCTCTCCC CTTTGTCCCC GCTTCTCCGC 17820
 CAGCGGAGCG TCCACACGCC GCGCTCGCTA TTTTCTATG GAGAGGTATG GAGAGGAGT 17880
 GACAGCGCTT CCGATCTCAT ATCTCGCGG ACGTCTCTG CTGCGCGGCG GCGAGCGGCT 17940
 ACTGCGACCG TCTGAGGCG CTCTGAGGCG GATCGAGCC GACCCGCGG GATCGAGCT 18000
 GCTACCTCTC CCGCGCTGCG CTCTGCGGCG CGACCCGAG GGAACCGGCT TGATGCTGCT 18060
 TCGATCTCTC CGCGAAGAC TTCCACCGGA TGCCCGGGT GGGCGGTTG GAGTCAAGT 18120
 GGACACCCCC GAACGGTCT GTTCTTGGGG GTGGGTTGAC GTACAGGTTG GACTGCGAGC 18180
 CCGACAGATT TAAAGGGTGC GTGGGTATG AAATGTACCC TAGGATGCC TCTTCTCTCT 18240
 CGGTTCTGCTC TCAGCTGCTC CAGGCGTGAA GACAACCTCC CATCGGAAC TCTTCTCTCT 18300
 CTTTCTTCCA CACACAGAT GAGACGCAAG AGAGGGAGAA ACAGCTCAAT AGATACCGT 18360
 GACCTTCAAT TGTGAATCT TCAGTCAATC ACACAAGAA CAGGTGACTA GCGAGGACA 18420
 CAGATCAAACT ACTATTTCGG GGTCTCTGTC GTGGGATTGG TCTCTCTCTC TCTCTCTCTC 18480
 TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC 18540
 TATCTAGTTC ACAGAGCACA CTCACTTCCC CTTTTCACAG TACACAGACT GAGTAAAGAG 18600
 CGCCCAACCC TCCACCCGTT GGTGACGAA ACCCTTCTCT TACAATTGAT GAAAGAGAT 18660
 ATCTGGGCGG GGCACGCTAG CTCACGCGCT TCACCTCCGG ACTTTGGGAG CCGGAGCGCG 18720
 GTGATCTGCT TGGGCGCGGG AGTTCTGAGC CAGGCTGGCC GACTTGGCGA AACCCGCTCT 18780
 CTTGAAAAA TAGAACGATT AGCCGGGCTT GTTGGCGTGG TCTTGAATC ACBACGCTCT 18840
 GGGAGACTGG GCGGGGCGAT TTGTTTCAAC CGGGGAGGCC GAGCCGCGCA TGAGCTGAGA 18900
 TGTGCGGCTG GCGATGCGG CTGGATGACG GAGCGAGACC CCGCTCTCGA AGAATCATGA 18960
 TGTATTATATA AGATGAGTGT TGGCGGCTGA TGGCGGCTGT TAGTCTGCGC TACTCTGGAG 19020

GCTGAGACGA GGAGAAGATC ACTTGAGGCC CCACAGGTGC AGGCTTCGGT CGGCCGTGAC 19080
 CCACTGTATC CTGGGCAGTC ACCGGTCAAG GAGATATGCC CTTTCCCCTG TTGCTTTTCT 19140
 TTCTCTCCCT TCTCTTTTCT TCTTTTGTCT TCTCTTTTCT TCTTTCTTCT TTTCTTTTCT 19200
 TTCTTTCTTT TTTCTTTTCT TTTCTTTTCT TCTCTTTTCT TCTTTCTTCT TCTCTTTTCT 19260
 TGCTCTTTCT TCTTTCTTCT TTTCTTTTCT TCTCTTTTCT TCTTTTCTCT TCTTTTCTCT 19320
 CTCCCAAAGT GCTGGGATGT TCTGGGATGT TCCCTCCCTT GCACCATGCC TTTCTTCCCT 19380
 TCTTTGAAAG TGAGAGCGAG AGAGCGGCCT CCAGTGTATCT CATGTACGTG TTTAGAGACCT 19440
 GCATCTCGCT CGCTCACCCC GGCAGTGGTG CCGATCGTAAC TCATCTCCCT GCTGCTGGAT 19500
 GCTCCTGGAC TCGAGCGATC CTTCACACCT AGCCTCCAGA GTACAGAGCC TTGGACCGCG 19560
 GGACCGCGCC ATTTGTCCTA CACCGTTTTT AATTGTTTTT TTTTCCCCTG AGACAGAGTT 19620
 TCATCTCTGT GGCTTAGACT GCAGTGCGGT GCGCGGATCT TGCTCCACCG CAACCTCTGT 19680
 TCTCCGGTTT CAAGCGATT TCTGTCACTG GCCTCTGAGT TAGCCCGGAT TGCGGGCATG 19740
 CGCTGCCACG CGAACTCGAT TTCTGATTTT TAGTGGAGAG GGGGCTTTCT CATCTCATCT 19800
 GGGCTGGTTT CGAATCAGGT ACCTCAGGTG ATCCGCCCTC CACCTCCGCT GTACTCATCT 19860
 GTGATGACAG CGTGAGCCAC CGCGCCCGGC CTTCATTTTT AAATGTTTCT CCACAGACGG 19920
 GGTCTCATCA TTCTTTTGCA ACCCTCCTGC CCGCGCTCTC GAGTGTCTGG CGTGACGGCG 19980
 GTGAGCCACT GCGCGTGGAC TCCGGGGAAT GACTCACGAC CACCTCCGCT CTACTGTATC 20040
 TTTCTTTCTT TTTCTTTCTT TTTCTTTCTT TTTCTTTCTT TTTCTTTCTT TTTCTTTCTT 20100
 TGAATTATCT TATGATTAT TTTGTACTT TTTTTCAGAC GAGAGTCTCG TCTGGCGGG 20160
 GCGAGGCGAG GCGAGTTCAT GCGCATGCT TTTGAAGCTG CGGCACGCC TTTCAAAGCC 20220
 CCAATTCTGT CACACAGGCG TTTATCCCTT CTTGGAGTTG CAGCTGATG CTTCCTGTAG 20280
 TCTGGGCTTC TTTCCATTGC GAGCTTTCAG AGGCGCAGGG CCACCCAGAG GCTGGCTGCG 20340
 CTGTAGGATT TTGGGGTGTG AAACGTGGTG CCTATTTTT TATACCTGAG 20400
 CCGACACATC CCGCGACGC CATCGCTTGC TCGCCCTCTG AGATCCCCG CTCTCCACCG 20460
 TCTTCAGGCT CCGCTCTTAC TTTCACTTCT TCTTTCTTCT CGTTTGAAGT GGGGGTGGG 20520
 GAATGAGGGT GTGTGTGGG AGGGGGTGGC GGGTGGGAG GAGAGGGAGG GTCTTAAAG 20580
 TCGATTAGT GTCATGCCCT TTTCAACACC ACCACACCA CCGAAGATGA CAGCAAGGAT 20640
 CGCTTAATA CCGCGTGTTC TCATCTAGAA GTGGGAATT ACAGATGACA GTTCTTGCAT 20700
 GGGCAAGC AGGGGGAGCG GGGAGCGCGA AGTCTGCTGT CCGAGGAGGG GGTGTGAGGA 20760
 GAGACACCT CAGCAAGAAA ACMAAACACG AATACTGTGC GACACAGCAC TGACTACCCG 20820
 GGTGATGAAA TCTGTCTGAC ACTGAACACC CCGTGCACAA GTTTACCTAT GTACAACTCT 20880
 TGCACATGTA TCGCTTGAAC GACAAATAAA AGTTAGGGGG GAGAAGAGAG GAGAGAGAGA 20940
 GAGAGAGAGA GACAGAGAGA GAGAGAGAG AGGAGAGAGG GAGAAACGAA 21000
 CACCACTCTC TTGACTGTAG TCAGGGGGTT TCTGGCCCTT TGGGAGAAGC TTTACGAGCA 21060
 ATGCAGIAT TTGGGCCGCTT CTTTTTTTTT TTTCTTTCTT TTTCTTTCTT TTTTGGACT 21120
 GAGTCTCTCT GCGTCTGTCA CCCAGGCTGC GGTCTCTTCT GCGCTCTTCT GGCTCACTGA 21180
 AACCTCTGCT TCCCGGGTTC CAGTGATTCT TCTTCGGTAG CTGGGATCAG AGGCGCACAC 21240
 CATGACGGCG GGCTCATATT CCTATTTTCA GTAGAGACGG GGTTTCTCCA CGTTGGCCAC 21300
 GCTGTGCTCG AACTCTGAC CTCAAATGAT CCGCTTCTCT GGGCCTCCCA CAGTCTGAGA 21360
 AACGACAGCG CTGAGCGCGG GGGATTTCAG CCTTTAAAAG CGCGGCCCTG CCACCTCTTC 21420
 CTGTGGCCCT TACGCTCAGA ATGACGTGTC TCTCTGGCG TAGGTTGACT CTTTGAAGTC 21480
 CTTAGGCCAT TGCATGTAG CTTGGGCAGC AAGAGCCAAA CTCGNNCCCT CCACCTCTCT 21540
 GCGCACATAA TAACATACTA ACAAACTAAC TAACATACTA AACTAACTAA TAACTAAATA 21600
 TTTCTACAGC TCACCCATAA GTGTGTGTTC CCGTGAGAGT GATTTCCTAG AATTGGTACT 21660
 GTACACTGAA CGCAGTGGCT CACGCTGTGC ATCCGAGGT CAGGAGTTGC AGACAGCCCT 21720
 GGGCAACGCT GTGAACCCCT GTCTCTACTG AAAATACGAA ATGGAGTCTG CGCCCGTGGG 21780
 CGAGGACACT GTAAACCCAG CTACTCGGGA GCGTGGGGTG GAAGAATTGC TAAGAGCTGG GCGACAGAGT 21840
 CAGGCGGAGG CTGCAGTGAC CCAAGATCGC AACACTGCAC TACAGCTCTG ACATACATA 21900
 GAGACCCGGT ATCAGTACAT ATACGTACAT ACATACATA ATCAGTACAT AAGAARAGAA 21960
 ATACATACAT ACATACATA GAAARAGGAA GAAARAGGAA TGTATGTCTA TGTCTCTCTG 22020
 AAAAGAAAGA GAAARAGGAA GAAARAGGAA TGTATGTCTA TGTCTCTCTG TGTCTCTCTG 22080
 TGTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22140
 TCTCTTTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22200
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22260
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22320
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22380
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22440
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22500
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22560
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22620
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22680
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22740
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22800
 TCTCTCTCT TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC TCTCTGTTC 22860
 GGGCCCTGCG TATGCCAGA AAGTGAGAAG TCCCGCTCTG GTGTCATGAT TCGGGAGGTC 22920
 GGAATCTAGA CAGGCGGGCC TTGCTGGGCT TCCCACTCG GTGTATGAT

GAGGCCGGGT	CCCCCGTTGG	ATGCGAGGGG	CATTTCAG	CTTTCTCTC	GGTCACGTGT	22980
GGCGTCGGTA	CTTCTCTTAT	TTCCCGGATA	AGCTCTCTGA	CTTCAACATA	AACGGCGTCC	23040
TAAGGTCGGA	TTTAGTGTC	TGCTCTTTTC	ACCGCCACCA	CCGAAGATGA	AAGCAAAGAT	23100
CGGCTAAATA	CCGCGTGTTC	TCATCTAGAA	GTGGGAACCT	ACAGATGACA	GTCTTTGCAT	23160
GGGCAGAACG	AGGGGGACCG	GGNACGCGGA	AGCCTGTGTT	AGGGRGGAGG	GTGGGAAGGA	23220
GAGACAGCTT	CAGGAAGAAA	ACAAACACG	AATACTGTCT	GACACAGCAC	TSACTACCGG	23280
GGTGATGAAA	TCATCTGCAC	ACTGAACACC	CCCGTACAAA	GTTTTACCTAT	GTACAGCTCT	23340
TGCTCATGTA	TGCTTGAACG	ACAAATAAAA	GTTCGGGGGG	GAGAGAGAGG	GAGAGAGAGA	23400
GAGAGACGGG	GAGAGAGGGG	GGAGAGGGGG	GGGGAGAGAG	GAGAGAGAGG	AGAGAGAGAG	23460
AGAGAGAGAG	AGAAAGAGAA	TAAACACCAA	CCACCACCTC	CTTGACCTGA	GTTCAGGGGTG	23520
TTCTGCGCTT	TTGGGAGAAC	GTTACAGCAG	AATGCAGTAT	TTGGGCGCGT	TCCTTTTTTTC	23580
TTTCTTTCTT	TTTCTTTCTT	TTTTTTTGGG	CTGAGTCTCT	CTCGCTCTGT	CACCCAGGCT	23640
GCGTTCGGGT	GGCGCTCTCT	CGGCTCACTG	AAACCTCTGC	TTCCCGGTTT	CCAGTGATTC	23700
TTCTTGGGTA	GCTGGGATTA	CAGGTGCGCA	CCATGACGGC	CGCTCATCG	TTCTATTTTT	23760
AGTAGAGACG	GGGTTTTCTC	ACGTTGGCCA	CGCTGGTCTC	GAATCTCTGA	CCACAATATG	23820
TCTACCTTCC	TGGGCTCTCC	AAAGTGTCTG	AAACGACAGC	CTTGAGCCCTG	CGGGATTTC	23880
GCTTTTAAAA	GGCGCGCTGC	CTGCCACCTT	TCGCTCGGCG	CTTTACGCTC	AGAAATGACG	23940
TCCTCTCTGT	CCATAGTGTG	ACTCTTTGAG	CCCGCTAGCC	CATTGCACCT	TAGCCTGGCG	24000
AGCAAGAGCC	AAATCTCGTG	CCGCCACCTC	CCCGCGACCA	TAATAACTAA	CTACTTAAC	24060
AACCTACCTA	AACTCTTCAG	CGTCAACCAT	AACTGTGTGT	TCGGTGTGAG	AGTGATTTCT	24120
AGAAATGTGT	ACTGTACACT	GAAAGCAGGC	TTACGCTCTG	TCATCCCGAG	GTCAGGAGTT	24180
CAGAGCCAGC	CCGCGCCACG	TGGTGAACCC	CCGCTCTCTA	CTGAAATAC	GAATTAAGAT	24240
CAGCGCGCGT	GGGCGACGCA	CTGTAAACCC	CAGCTACTCG	GAGGGCTGGG	TGGGAAGAA	24300
TGTTGTAACC	TGCAGCGCGT	AGGCTCCAG	GACCCAAAGT	CGCACCACTG	CACACAGACC	24360
TGGGCGCCAG	ACTGAGACCC	GGTCTCCAGA	TAAATACGTA	CATAAATAAA	TACACACATA	24420
CATACATAGA	TACATACAAC	ATACATACAT	ACAGATATAC	AGAAAGAAAA	AAAAAGAAAG	24480
AAAAGAAAGA	GAAATAAGAA	GAAAGGCAC	TGTAATGTCT	CTGGGCTAGG	CGCTTCTCTC	24540
TGTCGTCTTC	TCTCTGTCTG	TCTCTGTCTT	TCTCTGTCTG	TCTCTTCTCT	TGTCGTCTCT	24600
TCTGTCTGTC	TCTCTGTCTT	TTTCTTTCTT	TCGTCTCTCT	TCCTTGTCCC	TCTCTCTCCC	24660
TCTCTGCGCT	GTCTCACTGT	GTCTGTCTTC	TATCTTACTC	TCTTTCTCTC	CCGCTGTCTC	24720
TCTCTCTCAC	CTCCCTCCGT	TCGTCTTCTC	TCTCTCTCTC	TTTCTGTCTC	TTTCTGTCTC	24780
TCTCTGTCTG	CTCTCTCTCT	TCTCTATCTG	TCTCTTTCTC	TGTCGTCTCT	CCGCTCTCTT	24840
TCTTTTTCTG	TCTCTCTCTG	TCTGTCTCTC	TCTCTCTCTG	TGCTTACTCT	CTGTCTTACT	24900
CTCTTTCTCT	GGCTGTCTGT	CTGTCTCTCT	CTGTCTCTCT	CTCCCTTTCT	GTCTCTCTCT	24960
CTCTCTCTCT	CTCTNNNCCC	TCCCTGTCTG	TTTCTCTCTG	TCGCCCTCTC	TTTCTGTCTC	25020
TCTCTCACTG	TCTCTCTCTG	TCGTCTCTGT	TCATCTCTCT	TGTCCTCTCT	TCTGTCTCTC	25080
TCTCTCTCTG	TCTCTCTCTC	TCGTGTGTGA	TCCTTTGTCT	TACTCTCTCT	CTCTCTCTGT	25140
CCGCTGTCTC	GTCTGTCTCT	CTCTCTCCCT	GTCCCTCTCT	CTTTCTGTCT	GTCTCTCTCT	25200
CTCTCTCTCT	CTCTCTCTCT	CTGTCTCTGT	CTTCTCTCTG	CTGCTGCTCT	ATCTCTGTCT	25260
CTGCTCTCTT	CTTCTCTCTT	CTGTGTCTCT	CTGTCTCTCT	CTGTCTCTCT	CTGTCTCTCT	25320
TTACTCTCTT	TCTCTGCGCT	TCATCTCTCT	TCTCTCTCTC	TGTCCTCTCT	CTGCTCTCTC	25380
TGTTTCTCTC	TCTCTCCCTC	TCTCGCTCTC	TCGTCTTTTC	TCTCTTTCTC	TCGTCTTCTC	25440
TGTCCTCTCT	TGTCGGTCTC	TGCTTTTTTC	TGTCGTCTCT	TCTCTCTCTT	TCCTTCTCTC	25500
GTCTGTCTCT	GTCTCTGTCT	CTGTCTCTCT	CTCTCTCTCT	CTCCTTGTCT	CTCTCACTGT	25560
GTCTGTCTCT	TGCTCTTACT	TCTCTTCTCT	CTGCTCCATC	TGTCGTCTCT	TCTCTCTCTC	25620
TCTCTCCCTA	CCCTTCTGTT	TCTCTCTCTG	TAGCTCTCTC	TCTCTCTCTC	TGTCCTCTCT	25680
TTTCTCTCTC	TGCTTTTCTC	TGCTGTCTCT	TTTCTCTCTC	TGTCGTCTCT	CTGTCTCTCT	25740
TCTCTGTCTC	TGCTCTCTCT	TCTCTCTCTC	TCTCTCTCTC	TGCTCTCTCT	CTGTCTCTCT	25800
TCTCTGTCTC	TGCTCTCTCT	TCTCTCTCTG	CTCTCTCTCT	CTCTCTCTCT	CTGTCTCTCT	25860
CTCTCTCTCT	TGCTCTCTCT	CTGCTCTCTT	CTCTCTCTCT	CTCTCTCTCT	CTGTCTCTCT	25920
GTCTGCGCTT	CTCTTTCTCT	TCTGTGCTCT	CTCTCTCTCT	CTCTCTCTCT	CTGTCTCTCT	25980
ATCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	CTCTCTCTCT	TGCTCTCTCT	CTGTCTCTCT	26040
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	CTCTCTCTCT	TGCTCTCTCT	CTGTCTCTCT	26100
TTTCTCTCTG	TCTCTCTCTG	TCTCTCTCTG	CTCTCTCTCT	TGCTCTCTCT	CTGTCTCTCT	26160
TCTCTGTCTC	TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	CTGTCTCTCT	26220
TGTCGTCTCT	TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	CTGTCTCTCT	26280
CCCTCTCTCT	CCCTCTCTCT	CCCTCTCTCT	CCCTCTCTCT	CCCTCTCTCT	CTGTCTCTCT	26340
CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTGTCTCTCT	26400
CTGTCTCTCT	CTGTCTCTCT	CTGTCTCTCT	CTGTCTCTCT	CTGTCTCTCT	CTGTCTCTCT	26460
TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	TGCTCTCTCT	CTGTCTCTCT	26520
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	CTGTCTCTCT	26580
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	CTGTCTCTCT	26640
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	CTGTCTCTCT	26700
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	CTGTCTCTCT	26760
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	CTGTCTCTCT	26820

CACGTGTGCT	GTCTTCTGTC	TTACTTCTTT	TCTTTGCTCT	GCCTCTCTGT	CTGTCTGTCT	26880
CTCTCCCTCC	ATGCTCTCTT	CTCTCTCTCT	CTCTCTCTCT	CTCCGTCTCT	CTCTCTCTCT	26940
GTCGTGTTCT	CTCTCTGTCT	GTCTCTCTCC	CTCCATGTCT	CTCTCTCTCT	CTCTCTCTCT	27000
CTCTCTCTCC	GTCTCTCTCT	CTCTTTCTGT	CTGTTTCTGT	CTCTCTCTGT	CTCTCTCTGT	27060
CCATGCTCTC	CTCTCTCCCT	CTCACTCACT	CTCTCTCCGT	CTCTCTCTCT	CTCTCTCTCT	27120
GTTTCTTGCT	CTGTCTGTCT	GTCTGTCTGT	GTCTGTCTCT	CTCTCTCTCT	CTCTCTCTCT	27180
CTCTCTGTTT	GTCTTTCTCC	CTCCCTGTCT	GTCTGTCTGT	CTCTCTCTCT	CTCTCTCTCT	27240
CTCTGTCTCT	CTCTCTCTCT	CTTTCTGTCT	GTCTCTCTCT	ATCTCTCTGT	GTCTCTCTCT	27300
GTCTTTCTAT	GTCTGTCTCT	TTTCTCTGTA	GTCTGTCTGA	CACACCCGTG	CGGTGAGGCG	27360
CCTGCCCTTC	CACGAGAGTG	AGAAGCGCGT	GTCTCGGTGC	CTAGAGAGGC	CGAGAGGAGT	27420
CTAGACAGAG	GGGCTTGTCT	GGGCTTCCCC	ACTCGGTGTA	CGATTTCCGG	AGSTCGAGGC	27480
CGGGTCCCGC	CTTGGATGCG	AGGGGCATTT	TCAGACTTTT	CTCTCGGTCA	CGTGTGGCGT	27540
CCGTACTTCT	CCATTTCCCC	CGATAAGTCT	CTCTGACTTT	AACTCAAACT	GTAAAGCCCG	27600
GACGCCAACCA	CGCGAAACCT	CCGTCTCTAC	TAAAAATACA	AGCTCGACTC	GGAGCGGGTG	27660
GGGCGAGCGC	TGTAATGCCA	GCTCCTCGGG	AGGCTGAGGC	GGGAGAATCG	CTTGACACAG	27720
GGAGCGAGGG	GCTCGAGGGA	GCCGAGATCG	CGCCACTGCA	CTACGGCCCA	GGCTGTAGAG	27780
TGAGTGAGAG	TCCGTTCTCTA	AATAAATACG	GAAATTAATT	AATTCATATA	TTCTTTTCCC	27840
TGCTGACGGA	CATTTCGCTA	CAGGCATCGG	TTGCTTTCCG	GCATCACTCA	CGGGCCAGTG	27900
TATTGGAAG	TGCGACGTGA	CACGAGAGCA	GGTCTCGCCG	ACTTCAACGA	GCCTCGGGCA	27960
ACGGGTTTCT	CTCTCTCTCT	TCTCGAGGCC	CCTCCCTCTC	TCCCTCGTGT	CTAGGGGAAC	28020
CTCGGCTAGG	GAACTCTCCG	CTCGGGGCGC	CTATTGTTCT	TGATGTGGCG	CTTTACTTTT	28080
CTTTGTGTTT	TGCGGCTCTG	ACTCTTCTAC	TTGGGCTTTG	GGAGGCGTCA	GTTTAATTTT	28140
TAGGTTGGGT	TTCCGCTCCG	CCCACTACCC	ACGTCCCTTC	ACCTTAATT	AGTGAGNCGG	28200
CTTCCAGAG	CTCCGCCCAA	CCGCCCCCCC	CCCCCGCGCT	CCCAACACCC	TGCTTGGAAA	28260
CGATCTCAT	CTGCCCAGGC	CGCCCCGGGT	GTGCTCCGCT	CTTCTCTCCC	CCCTTGCCTG	28320
CGACGTTTCT	GAGTGAAGGG	TGACATTTGC	ATCGGTGGGC	GTACGGCTCT	ACTCTGGGAGC	28380
GGCATAGCCC	CTTGACCCCG	GTGGGCAAGC	GGGCGGGTCT	GCAGTTGTGA	GGGCTTTCCG	28440
CCCGCTGCTT	CCCGCTCAGG	CTCCCTCCCC	TAGGAAAGCT	TCACTCTGGC	TGCGGCTCGG	28500
TACACTTTTA	TCAGATTTAT	TTAGTTTCTC	CGCCCTCCGG	CCAGCAGAGT	TTTCAAAATG	28560
GAAGGCGCGC	ACCGCTCTAG	TCTGGGCTTT	CTCAGACTTT	GCCCAAATAA	GAAACGCTTT	28620
CTGAAAAATC	ATAACTTTTC	TCACTTAAGA	TTTCCAGGGA	CGGCGCGCTG	GCCCGTGTTT	28680
GTGTCGTTGT	TTTGTGTTGT	TCTGTTTGTG	TTTTTCCITT	TTTTTCCITT	TCGTATGTGC	28740
TTTCTTTTCA	GGTGAAGTAG	AAATCCCCAG	TTTTTCAGGA	GACGTCTATT	TTCCCCAGTA	28800
CACGTTAGCT	CGCGTTTTTT	CTGTTGTGTA	ACTAGCGCTT	TTGTGACTCT	CTCAACGCTG	28860
CAGTGAGAGC	CGGTTGATGT	TACNATCTCT	TCATCATGAC	ATCTTATTTT	CTAGAAATCTG	28920
GATAGCGAAT	GCTGCTGACT	CTCTTGTGTC	TGTTGTGTTT	GTGTTGTGTT	TGCTGCTTGC	28980
TGTTGTGCTT	GTGCTGTGTT	TGTGCGTTGT	CGTTGTTTTC	AAAGTATACC	CCGGGCCACG	29040
TTTATGGGAT	CAAAAAGCATT	ATAAAATATG	TGTTGATTAAT	TCCTGAGCAC	GGCCTTCCTC	29100
CCCGCTCTCT	TGCTCTCTCT	CTCTGCTCTG	TCTCTCTCTT	TCCTGTGCTG	TCTTCTCTCT	29160
CTCTCTCTCT	CTGTCGCTCT	CTCTCTCTCG	CTGTCGCTCT	CTCTCTCTCT	GGCTCTCTCT	29220
CTCTCTCTCT	CTCTGACTGT	CTCTCTCACT	GTGTCGCTCT	TCTGCTCTAC	TCCCTTTCTC	29280
TGTCGTCTGT	TGCGTCTCTC	TCTCTCTCTC	TCCCTGTCTG	TATGTTTCTC	TCGTGCTCTG	29340
TCTCTCTCTC	TCTTTCTGTT	TCTCTCTCTC	CGCTCTCTGT	TTTCTCTGAC	GTATCTCTCT	29400
TTTCTCTCTC	TCTGCTCTCT	TCTGCTGCTC	TCTCTCACIC	TGTCCTCTGT	GTATCTCTCT	29460
TCTCTGCTGT	CTGTCCTCTC	TCACTCTCTC	TCTCTGCTGT	TCTCTCTCTC	TCTTTCTGTT	29520
TCTCTCTGTC	TCTCTGTCCT	TCTCTGTCCT	TCTCTGTCCT	TCTCTGTCCT	TGTCGTCGTT	29580
TGTCCTTCTC	TCTCTCTGTC	TCTCTCTGTC	TCTCTCTGTC	TGTCCTCTGT	CTAGCTCTCT	29640
CTCTCTCTCT	CTCCCTGTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	29700
CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	29760
CTCTCTGCTG	CTCTGCTCTC	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	29820
CTCTGCTGCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	29880
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	29940
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30000
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30060
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30120
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30180
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30240
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30300
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30360
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30420
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30480
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30540
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30600
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30660
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	30720

CCGAGCCCTG TCGCTCCTGT CTCAAATATG TACGTGCAAA CACTTCTCTC CATTTCACAC 30780
 ACTACCACAG GCCCCTTGTT GAACCACTGG CTCTTTGAAA AAAATCCGAC AAGTGTGTTT 30840
 GGCTTTTGTG CTAGGAGGCC TAAGCCTGCT GAGAACTTTC CTGCCGAGAC TCCTCGGGAC 30900
 CATGCTTGCT AGCGCTGGAT GAGTCTCTGG AAGGACTCGC GGGACTCGGC AAAGCTGACC 30960
 TGTGCCACCG AGGTCAAATG GATACCTCTG CATTTGCCCG AGGCTCTCGA AGTACATCAC 31020
 CGTCAACAAAC CGTCAACGTC AGCATCTCTG TGAGCTCTGC CAAGCGGGCG CTCTCGGGCCG 31080
 GAGCTTCTGG AGCCCGGCCCT TCGTCCGCTA AAGTGCCAAAG GGATGGTGAC TTCCACCCAC 31140
 AAGGCTCCAC TGAAGCGCGA AGATGTGGAG CTGAGGTGAG AGAGGGGAGC ATGGAGGGAG 31200
 ACGTGCCGAC AGGCGAGCAG TTCCCAAGGC CTTCGGCCACC CCACCCAGCG CCCACGCCGCC 31260
 ACGTCCCGGG ACACCGCGCG ACACCGCGCG TTTATCCCTC CACTGTGTCA CCGCGCGGCC 31320
 CACCCCAACA CGCAACCCAC GCACACACGC TGGAGGTTCC AAAACCCACG GGTGTGACTA 31380
 GAGCTGACG GACGAGAGAG CCATTTCCAG AGGTGGGGTG GGTGGGGGTG GGGTGGGGTTG 31440
 GGGTGTGTGG GGTGTGTGGC GAGGCCGATT CTCCCTCTAG GGTGCTGACA GTGTGAGAAT 31500
 GAAATATGCT TCTTGAGGCG AGGCGCTTAC TTACGGCATC ACAGCTTCTGC GAGCATCTGC 31560
 TCAAACCTCT CAAATAAGAG ACAGAACCTG GACCTGACCC GTCCGAATTA ACACCTTGAC 31620
 GTGCTGGATG TATCTCTGTA TCTTCTGAGT GAGGTCTTGC TCTGTCCCC AGGCTGGAGT 31680
 TGCTTTTGTG TGTTTTGTGT TTTCTGAGAT GGAGTCTTGC GGGTTCAAGT GATTCCTGT 31740
 CGAGTGGCGT GATCTCAGCT CACTGGAACC TCTGCCCTCT GGGTTCAGT GATTCCTGT 31800
 TCTCAGCGCT ACCATCGGCG GCTCATTTTT TTTTTTTTTT TTTTGTGAG ACAGCTGGGT 31860
 TCACCCCTTT TCAATTTGTT TCACTGGAGA TTCTAGATTG GAGCCACACC TCMTTCGCT 31920
 CCACAGAGAG ACTTCTTTTT TTTTTTTTTT TTTTAAAGG CAACGCAACA TGTCTCGCT 31980
 ATTTGAGTGG TCTTCTTATAT CATTTAATAA TGTTTATAGA TGAAGAAAGC GTATTAAGA 32040
 CTGTGCTTAAT GTATGTTGAA GTGAGACAA AAGAAAGGCT ATCTAATTTG TGGTTAGAAT 32100
 AAGTGTGCTC AGTATTTAGA AGCTACCTAA ATACGTCAGC ATTTACACTC TTCTCTGATA 32160
 AAGCTGGCGC ATCTGAATAA TCCTCCTTTA AACAAACACA ATTTTGTATA GGGTTAAGAT 32220
 TTTTTTAAAG ATGCGACTCC TGCAAAATAG CTGAACAGAC GATACACATT TAAAAAATA 32280
 ACAACACAAG GATCAACACG ACTTGGGAAA AATTCGAAAA CCACACAAGT CTTTATAGA 32340
 ACTGAGTTCT TAAATATAGG CGGAGAACGT AGCTATCGGA AGAGAAAGCA GTATTGCAAC 32400
 GTTGATTGCT ACGTTGGTCA GCAGTAGCTG GCACATCTCT TTTGGCCATC TTTTCGGCAA 32460
 TGTAAATAGT ACAGAAAAAT GAGATATGAT CCATTAACAA ACATATTCGC AATCAAAA 32520
 GTGTTTCACT AATATAATGC TTCAGATTAA GAAGCAATC AATAGTAGA ACTCCACTGC 32580
 TGTAAATAGT CACCCCAAAG ATCCCGTAT CTGACAAAAT AACTACCACA GGGTTATGAC 32640
 TTCGAATCTA TACTTTCTTC TTGATATTTA CTATGTATT TATTTTTTTT AATTTATTTC 32700
 TCTTGAGACG CGTCTCGCTC TGTGCCGAG GCTGGAGTGC GATGGTGTGA TCTCGGCTCA 32760
 CTGCAACCGC CACCTCCCTG GGTTCAGAGC ATTCCTCTGC CTCAGCTCC CGAGTAGCTG 32820
 GGACTACAGG TGCCCGGCCAC CACGCCGAGC TAATCTTTAT ACTTTTAATA GAGACGGGT 32880
 TTCAACGTTG CGGCCGAGT GGTCTCGATC TCTTGACCTC TGACCCCGCC CGCCTCGGCC 32940
 TCCCAAAAGT CTGGGATGAC AGGCGTGAGC CACTGAGCCC GGCCTTCTCT TGACGTTTAA 33000
 ACTATGAAGT CAGTCCAGAG AAACGCAATA AATGTCAACG GTGAGGATGG TGTGAGGGA 33060
 GAGTAGAGAC CAGCTTTTTT CCTATCTTAT TCAGTTGATA ACAATATGAC CTAGTAGTA 33120
 ATTCTCTATG TGCTCTATTA TACACGAGTA CAAGAAGAGT AACACAGAGT ACTGCTAAT 33180
 TAAAGGTATC GTGAAGTTCT TCATAGTAAC TCCCTAAGCT GBAACACTGT CAAAAGCAG 33240
 CAGCTAGTGA ATTTGTTCCA TGTATTTTTT TATTATCCAA TAAGTGAAT ATGCTATTC 33300
 TTTCCAGTCT CCAACAGACT TCTGTGCCCT ATCAACACT CGGTGCTCGA AGAAAAGAT 33360
 AGCAAAATCA GGAACACAAG CTAAAGAAAC ACACACACAA ACCAAAGACA ACTACACGCT 33420
 CTGCAAAAGT TTGCTAGAAG ACTGAAACTG TTGAGTATAA GGATCTGTGA TTTACAGAT 33480
 ATGAGTTTCA TTGAGAGTTT GTTCAAGACA TACGTTTCGT AAGGAAACAT CTTAGTTAGA 33540
 AGTTATTCAG CAGTAGGTAT CATCCCTAAG TATTTTTTCA CAACTCGGTG ACAATAGAAG 33600
 GCTATCTAAC CAGAAAAAAT AGCGAGTAGC GGCACCATCC ATAGGCTTCT GTCTTTACGC 33660
 TTTATTGACA CTTTACATGC CTTACAATGT CTAGGATTGA CCTGATAGC ATTTTCGAAA 33720
 CAAGCTAATG CTTTGTCCAG TTCTTCACTG AAGACAATC ACGCCCTAAT CGCTATAGG 33780
 CATAGCATC TTTTGGATCC ACTTCGAGAG TTCTCTGAA GAATTGAATC GCAATATCGT 33840
 GTTCCCGTTT CGAGACCGAA ACAGTTTCCC TGCAGCACAC CAGGCCCTGT GCTGGCGGAT 33900
 TTTTATCATG GTCTGTGAAG TCTTTGGACA GAACTGAAAG AGCAACCTCT TFCGAGGAT 33960
 GCCAAAGTGT GTGAGAGTAG ATCTCCATGC CTTCGACTCT GTAATTCTCA ATCCCTCTAA 34020
 CCTCTGAGAA TTGCTTTTCA GCTTGCCTGG ACTCTGAAAG TTTAACAATAG CCGNTTTCCG 34080
 ATTTGGSCAA GTACCCCAAG GGTATTGAG TGGTGAAGAG CTAGATGGCT CAAAGTGTCT 34140
 ATAGTCTTCT TGCCGTGATA AGAACACAAA GCTAAATAAC CTTCCTCCCT CTAAGTAGA 34200
 AGGCTCATCA AGCCTTCCGC TGCTGCTTTT TGTAGATTAA AAGCCTGAAT CTGAGGCGCG 34260
 AINTGCGGTCA TTTTCCCTTC TGAATGAGC GAAGAGTCCA ATTTTGTGAT TTTCCAGCTA 34320
 TCACTTATGT TCGGTGGAGT TATTGCTCCT TTTATAGTTT TACTTTTGGT TCTCTGTTT 34380
 GGGATTTTAG GTGAAACCTT CATTTTTAAT TTTCTCTTAA TTTCTCTCGG GTTGTGAGCT 34440
 GTGCTAGTCT TACGTCGTGT AATTTCCTTG AGGNGCGTGC ATTTGGGGGA GTTGGCCATG 34500
 TGGGCTCTCA TACCTGAGGT GTTGCCCTTG TCGGCGGACC AGAAGCTTGT GTTTTTCGAA 34560
 GGACTGGAGT TACCTTTCCG CTCTTCCCC TCTCGAGAA GACAGACGCT GTTCCGGTTT 34620

GGCCGATTCT GGCAACAGCG TTTCTGAG GGGCTCCGGT GGATGGCAGC TCAGTGACAG 34680
ACGGTGTCTC ATACCACTGC AGTTTTGTCA ATAGGGTCCG TCTCCGGGAC TTGGGGTTTC 34740
TAATGGCAACA ATGCCAACAC TTGGGGTTTAA TGGACCTAACA GCTGCTGGATC CTCCTAATAA 34800
ACTTCGACCA GTTTTTGGTT TATGTTGAAC CTGTTTAGAT CATATGGAAG TTCTCTGTTC 34860
CAGTGGGACA GTATCAGGTG AAGGAGCAGC TGAATCGGATA GAAGACACCTG GGGACTCTGT 34920
ATTTCAAGGAG TACTTTGAAT TGGAGATATC TAAATTCATC CGTTTTCATT CGACGGTGTCT 34980
CTGGGGTGTG TCCGTAAAGT CGGTCTCGGG CTGTCTGTGA CTCTCTAGAG GACGAGGTCC 35040
AAGTGTGTGT GCGCAACACT TGGACAGGCA GTGCTATAAG CTCTCTAAGG AGGTGAATCA 35100
AAATGTTTGG TCAGAGTCTG GCITTTTCCC CCTATTTCAC ATCAATGAFTC AAAGGGACAC 35160
CAGAGGAAGG GATTTCACAG AAGGCTCTTT TGGTCACATT AGCAGAGCTGA ATCAACCAAG 35220
TCTGTCTTGG TAATATACAT TCCCGACGAT TGGTATAGAA ACACCTCCAC CGATAGGATT 35340
TCAGGAACGA AATATATCAT GTGGCTTTTC AGCAACACT TCCTGACAGC CATATTGCAT 35400
TGATCCCCCT TCTCAACTGT GCTGAGATAT AGCTATATAT TTTTATAGGC CTTCCTCTAG 35460
TCGGGGGTAG TCCCTTCAGG TCCCTTCCTG AGTGTACTTC TCATATAAAG 35520
CGGTAAAGAC AGGTGCGCAG TAAGAACAAG ATCTCGGTAG GCATAGTGGT TTAGTGTCTG CATATAGCA 35580
CGTTCTCGCG GGGAAACCGC ATCTCGGTAG TCTGAGGCTC AGGCCACTT TGTGCAAGTGC 35640
GCCTGGACGAG CCCCCTGCGC CACCGCTGAC CCGGCGCGGC CAGCCCCGGC TCATTAAAG 35700
CAGAGGACCC CGGTACCGCG CGATGCGGAG AGTCCGAGAC AGAATGACTT CTTTATCTGT 35760
CGACCCAGCCG CGCTTGTGAT CCATGCAAAA CCGAGAGTCA CTCCTGTGTT 35820
CTGACTCTGG TCCTCTCCA AGTTCAGTGG GGGATGTGGA AGGTTGAGAG 35880
AGGAAGCTCT TCCTCTCTCA TTCCGTCTGC ACCTGTGGGC TGAAGGCTGG AGCTCTCTCT 35940
CTTGGGAGAT CTGACTCTCT TACTCTCTCT TCCCTCTGGC CCACCCCTGG AACGCGGGC 36000
CGATGCCCCG GAGTGAAGCT GGCTCGCAC TCTCGGAAA GCTTGGAAA TCTCTCTACT 36060
CGCGCGGCCA CCGTGAARAC CAGGAGCTCG GGACACAGCT GCTTTCTGGA GAGAATGCTG 36120
AGAGTCTCTC CGCACTCTC TCTTGACTTG AGTTCCTGCT GGGTGTGCTGG TTAGAGACTGA 36180
GTGAGACGAC ATGTATTAACT TCAGGCGGGG TGCTGTGTGC TCACGCTGCT AGACCCCAAC 36240
CTTTGTGAGG CGGAGGCCGT AGGATCCCTC GAGGAATCGC CTAACCTTGG GAGGTTGTAG 36300
GTTTGCAGTGA GTGAGCCATA GTTGTGTCAC TGTGCTCCAG TTTGGGCGGA AGACAGAATG 36360
AGGCCCTGCC ACAGCGAGCG AGGCAGGCG GCAGGCGAGC AGACAACAGC TGATTATTGT 36420
TCTCTCTCAGG GTAGGAAGCA AAAATAACAG AATACAGCAC TTAATTAATT TTTTTTTTTT 36480
CCTTCGGAGC GAGTTTCACT CTTGGTGCCC ACGCTGGAGT CGAGTGGGAC CATCTCGGCT 36540
CACCCTAACG TCCACTCTCC CGCTTCAAGC GATTCCTCTG CCTCAGCCTC CTGAGTAGCT 36600
GGGATTAACG GGAGAGGCCA CCACACCCAG CTGATTTTGT ATTTGTAGTA GAGACGGGAT 36660
TTCTCCATGT TGGTCAAGCT GGTCTGGAAC TGGCGACCCC AGTGGATCTG CTGCGCCGCG 36720
CCTCCCAAAG TGTGCGGTG ACAGGCGTGA GCCATCGTGA CTGGCCGGCT ACGTTTATT 36780
ATTTATTTT TTAATTAATT TACTTTTTT TAGTTTTCCA TTTTAACTTA TTTTATTATT 36840
TACATTTATT TATTTATTTA TTTTATTACT TAITTTATTA GATCTCGGCT CACTGCAAGC GCGCCCGCCA 36900
CTGCTGCGCC GAGTCTGAGT GCACGCGCGT CCTCAGCCTC CCAAGTAGCT TTTTCTAGAG TGTGCGGATG 36960
GGGTTCAAGC CATTCTCTGT GTATTTTGTG TAGAGATCGG GTTCTACTGT GGTAGCAGG 37020
CCGTGCCCCG TCTCTCTGAC CCGTGATCCG CTCCACCTCG TATCTATTTA TTAACTTTGA GTCCAGGTTA 37080
ATGTTCTCGA ACAGGCGTGA TGAAACCACT TAGTTTTTGT AATTTTTTTT TTTTGTGAGA GTGAGGTTTCA 37140
ACAGGCGTGA AAGGCTTGGA CCGAGGGATC CCGCGCCCTT CCGCCTCCCA AAGTGTGCG 37200
GATGACAGCG GCGAGCCTAC CCGCGCCGGA CCGCCCTCTT CCGCTTCCCG CCGTGTCTG 37260
CCCGACAGAC AGTITTCACGG CAGGCGGTTT GGCTGCGCTG CTAAACTCCT CCGTGTCTG 37320
AAATTTGGGA CGTCACTCTC TGACCTCACG GCGAGCTGAG CAGCTCAAG AGTTCCTG TCTCAATAGT 37380
CTATCAGAGC ACGGTACAGG TAAGGAGGAG AAAAATCGTA AACAACAGAA CAGTCTCTT AGAATATGCC 37440
TGTGATACAG AAATACACGG ATTCACCCAA GTGATTTCTT TTTGCTTGG AGCTTACTG AGAGATTGCC 37500
TTAGCCCTGG TGTGCGTGCC AGTGATTTCT GACGGAAGT GTTGTGGGA GTTGTGGGA CCGCTCCGAG 37560
CAGTGCCTCT CTGCTCTCTG TGGTCTGATT CTGCTGCTT CCCCACACC CTCTCCGAG GGGATCCCA 37620
CTGCGTCCCG CTGAGGAGCC CCGTGGGAGC GCTGAGGAGT CCGTGGGAGC GTACCTGAG GTACCTGTT 37680
TTCTCTCGCG GCGAGGCTCA TCACTGGCAG CAGAGGAAGG GAGCACAGG CCGCTCTGCG ACACCTCGG 37740
CTCTCTCACG CTTCTCACCG GCCCGGCC CGCTCCAA GTTCTCTCCT CCTTGTGTC 37800
GTAGGAATTC GACTCTTGA CGACCGGGCT TGATAGACT TTGATCAGG AAAACGAAC 37860
AATCAGATTA ATAAATATA TAACACAAA GTAACCTAAT AAATAAAATA AGTCAATAC 37920
ACCCATTACA ATACATAAAG ATACGATACG ATAGAGTGGC ATAGGATAGC ATAGGATACA 37980
ATACAATAGG ATACGATACA ATACATATCA ATACAATACA ATACAATACA ATACAATACA 38040
ATACAATAGC ATACAATACG CCGGCGCGCG TGGCTCATGC CTGTATCCG GTCACTTTGG 38100
GATGCGGAGG TGGAGCGATC ACCTGAAGTC GGGAGTTGGA GACAAGCCCG AACCAATGCT 38160
AGAAATCCCG TCTCAATTGA AAATACAAA CTAGCCGGGC GCGGTGCGAC ATGCCATAAA 38220

TCCGAGCTGC	TAGGAAGGCT	GAGGCAGGAG	AATCGCTTGA	ACCTGGGAGG	CGGAGGTTGC	38580
AGTGAGCGGA	GATTGCGCCA	TCGCACTCCA	GCTGTAGCAA	CAAGAGCGAA	ACTCCGCTCT	38640
AAAAATAAAT	ACATAAATAA	ATACATACAT	ACATACATAC	ATACATACAT	ACATACATAC	38700
ATAAATTAAT	ATAAATAAAT	AAAATAAAT	AAATAAATGG	GCCTTGGCGG	TGGGCTCAAG	38760
CTGTCTCATCC	CCTCACTTTT	GGAGGCCAAG	GCCGCTGGAT	CAAGAGGCGG	TGAGCAACAA	38820
AGGGCCAGTA	TGGTGAATCC	CCGTCTCTAC	TCACAATACA	CAACATATAG	CGGGCGCTGT	38880
GCTGTGCTGT	ACTGTCTGTA	ATCCAGCTTA	CTCGGGAGGC	CGAGCTGAGG	CAGGAGAGAT	38940
GCTTGAACCT	GGGAGGCGGA	GTTTGCAGTG	AGCCGAGATC	GCGCCACTGC	AACCCGACCT	39000
GGGCGACAGA	GCGAGACTCT	GCTTCCAAAA	AATGAAATAT	AAAATGAAAC	GCACACAAAT	39060
AAATTAATAAG	TGAGTTTCTG	GGGAAAAAGA	AGAAAAAGAA	AAAGAAAAAA	ACACACAAAC	39120
AGAACAAACC	CACCGTGACA	TACACGTACG	CTTCTCGCCT	TTTCAGGCGCT	CAACACAGTT	39180
AGGAMTTATG	CGTGATTCTT	TTTTTTAACT	GTATTTTATG	TTATTATCAT	GATTGATGTT	39240
TCGAGACAGGA	GTCTCGGAGG	CCCGCCCTCC	CTGGTTGCCC	AGACAAACCC	GGGAGACAGA	39300
CCCTGGCTGG	CCCGGATATG	TCTTCTCTTT	GGTCAGGGGT	TTCCCTTGCT	TTCTCTTGCT	39360
CTTTAACCCG	CGTGAGACTC	TCCGCCCTCG	GTTTGACAGA	TGGCAGCTCC	ACTTTAGGCC	39420
TTGTTGTTGT	TGGGGAGCTT	CCTGATCTCT	CCAGATGTGA	TGGAAGAGCAG	GTAGATTGCC	39480
TTGCCCTGCC	TTGCTCGGCC	TTGCTTTTTC	TTTCTTTCTT	TTCTTTCTTA	TGATCTTTCT	39540
TTTCTTTCTT	TTCTTTCTTT	TTTCTTTTTC	AGACAGAGTT	TCACTCTCTT	TGCCCGAGCT	39600
AGAGGCGAAT	GCGCGGATCT	GCGCTCACCG	CACCTCTCCG	CTCCAGAGTT	CAAGCGATCT	39660
TCCTGCTCCA	GCTCTCTGAT	TAGCTCGATG	TACAGCGATG	GGCCACCGAT	TCGCTCTGAT	39720
TTTGTACTCT	TAGTAGAGAC	GCTGTTTTTC	CATGTGTGTC	AGGCTGGTCT	CCCACTCCCA	39780
ACCTCAGAGT	TCGCGCTGCG	CTTAGCCTCC	CAAAGTGTCT	GGATGACAGG	CGTGCAACCG	39840
CGCCACGCTC	CTCTCTCTCT	CTCTCTCTCT	CTCGCTCGCT	TGCTGTCTGT	CTTTCGCTGT	39900
TTCTTGCTTT	CCGCTTTTCT	TGCTTTCTTT	CTTTCTTTCT	TTTCTTTTCT	GCTTGTCTTT	39960
TTGCTTGCTT	CTGCTTTTCT	GTTGCTTTCT	CGTTTCCGTT	TTTCTTTTCT	TTCTTTTCTT	40020
TTTCTTTTCT	TGTTTCTTCT	CTGCTTGTCT	TTCTTGCTPT	CTTGTCTTGT	TTGTGCTTGT	40080
CTTGCTTTCT	TTTCTTTTCT	CTTCTTTTCT	TTCTTTCTGT	CTTCTTTCTG	TTTCTTTCTT	40140
GCTTGCTTGT	TTTGTGCTTT	TCTTGTTTTC	TCGATTTCTT	TTCTTTTCTT	TTTCTTTTCT	40200
TGCTTGCTTT	CTTGTGCTTT	TGCTTTGCTG	CTTCTGCTGT	TTCTTTTCTT	TTTCTTTTCT	40260
TTTCTTTTCT	TTTCTTTTCT	TGCTTGTCTT	CTTGTGCTGT	TGCTTTGCTG	CTTGTGCTGT	40320
TCGATTTTCT	TTTCTTTTCT	TGTTTCTTTT	CTTGTGCTGT	TTTGTGCTGT	ATTGCTTTGT	40380
TGCTTTTCTT	CTTCTTTTGT	TTTCTTTTCT	CTTGTGCTGT	TTTCTTTTCT	GCTTCTTTGT	40440
TTTCTTTTGT	CTTGTGCTGT	TGCTTTTGTG	CTTGTGCTGT	TTTCTTTTGT	TTTCTTTTGT	40500
GTTCTTTTCT	TGCTTTTGTG	CTTGTGCTGT	TGTTTCTTGT	CTTCTTTTGT	TGCTTTTGTG	40560
CGTGCTTTCT	TTTCTTTTCT	CTTCTTTTCT	TTTCTTTTCT	TTTCTTTTCT	TTTCTTTTCT	40620
CTTCTTTTCT	ATCATCATCT	TTTCTTTTCT	CTTCTTTTCT	TTTCTTTTCT	CTCATCTTCT	40680
TTTCTTTTCT	TTTCTTTTCT	TTTCTTTTCT	TTTCTTTTCT	TGCTTCTTGT	GAGACAGAGT	40740
TTTCTTTTCT	TTTCTTTTCT	TAGAGTGCAA	TGGCGGATCT	TTGGCTCACC	GCACCTCCGC	40800
CTTCCGCGGT	TCCTCGCTCT	CTCTCGCTCT	CAGCCTCCCG	ATTAGCGGGG	ATTAGCAGGG	40860
AGGCGACCCC	AGCGCTGCTG	TGGCTGATGT	TTGTGTTTTT	AGTAGGCAAC	CGGTGCTCTCT	40920
CCATGTTGCT	CAGGCTGGTC	TCCAACCTCC	GACCTCCTGT	GATGCGGCCA	CTCGGGCTCT	40980
TCGAAGTGCT	GGGATGACGG	GCGTGACGAC	CGTGCCCGGC	CTGTTGACTC	ATTTGCGTTT	41040
TTTATTTCTT	TGCTTTCCAC	GCGTTTACTT	ATATGTATTA	ATGTAAACGT	TTCTGTACCG	41100
TTTATATGCA	ACAAGCACAA	CGTGTATCTC	TGCAATGAAT	ACTCTTGGCT	ATGTTGAAAT	41160
CGATCGGTTT	GATTCGATAT	AGACTTCTGT	ATGATGATAT	TAGGTTGCTG	TGTTATACAA	41220
ATAATACACG	ATGCTCTTAT	AAGACTGATG	TCAGCTGATA	AGACGTTGTA	TTTACAGTTA	41280
AAAGCGCTCG	TATTTATGTT	TGTAATGAA	CCGAGCGTAC	GTAGTTATCT	CTGTGTTTCT	41340
TTTCTCTCTC	CTTGTGTTTT	TTTCTCTCTC	CTTCTCTCTC	TTTCTCTCTC	CTTCTCTCTC	41400
TTTCTCTCTC	CTTCTCTCTC	TTTCTCTCTC	CTTCTCTCTC	TTTCTCTCTC	CTTCTCTCTC	41460
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41520
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41580
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41640
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41700
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41760
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41820
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41880
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	41940
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42000
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42060
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42120
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42180
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42240
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42300
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42360
TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	TTTCTCTCTC	42420

AGGCTGGGGA	CGCCCTTCCC	GGCCCGGTG	CGGTCCGCTC	ATCCTGGCCG	TCTGAGSCGG	42480
CGGCCGAATT	CGTTTCCGAG	ATCCCCGTGG	GGAGCCGGGG	ACCGTCCCGC	CCCGGTCCCC	42540
CGGCTGCCCG	GGAGCGGTCC	CCGGGCCGGG	CCGCGGTCCC	TCTGCCCGCA	TCCTTTCTGG	42600
CGAGTCCCGG	TGCCCAGTCG	GAGAGCGCTC	CCTGAGCCGG	TGCGCCCGCA	GAGTCTCGCG	42660
TGGCCCGCCG	TCTGGTCCCTC	GTGTGTCGCG	GTCTGAGGAG	GGCCCGCGCG	AAATGCTTTC	42720
CGCTCCCGC	CTTGGAGACA	CGGCCCGGCC	CCTGCGTGTG	GCCAGGCGCG	CCGGGAGGCG	42780
TCCCCGGCCC	GGCGCTGTCC	CCGCGTGTGT	CCTTGGGTGT	ACCAGAGGGA	CCCGGCGCGC	42840
TCCGTGTGTG	GCTGCCATGG	TGGCGTTTTC	GGGACAGGT	GTCCGTGTCT	GTGTCCGCGC	42900
TCCGCTGGGC	CGGCGCGGTG	GTCCGTGACG	CGACCTCCCG	GCCCCGGGGG	AGGTATATCT	42960
TTCGCTCCGA	TTCGGCAATT	TTCGGCCGCC	GGGTTATAT			42999

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTCCCCGGCG	GCCCCGTGT	TCGCCGTTC	CGTGGCGCGG	ACAATGCGGT	TGTGCTCCA	60
CCTGTGCGTG	TCCGTGCAGT	GCCGTTGTGG	AGTGCTCGC	TCTCTCTCTC	CTCCCCGGCA	120
GCGTCTCCAC	GGTTGGGAGC	CACCGGTGAC	CTGCGCTCT	TCGGGCTTGG	ATCCG	175

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 755 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGTCTGGTGG	GAATTGTTGA	CCTCGCTCTC	GGGTGCGGCG	TTTGGGGAAC	GGCGGGGTG	60
GTCTGCTCCG	CGCGCCGAGC	TGTGTCCGGG	CCCACTTCCC	GCTCGAGGGT	GGCGGTGCGG	120
GCGGCGTTGG	TAGTCTCCCG	TGTTGCGTCT	TCCCGGCTCG	TTGGGGGGGG	TCCCGTCTGT	180
TTCCGGGGCG	CGGTGTCTTG	GCTTACGCTG	CCTGTGTTTG	GGACTGCTCT	AGGAGTCTGT	240
GCGCGGTGTA	TTCCCGCCGG	TTTTGCTCTG	GCTCTGCTCG	CTTGTGCTCG	GGTTTGTCTG	300
GTTCGTGCTC	CGGAGAGCGT	GGTTTTTTTT	TTTTTCGGGT	CCCGGGAGA	GGGTTTTC	360
CGGGGAGCGT	TCCCGTGGCC	CCTTGCAGCC	GGTGGGTTTC	CGTTTCGGGC	TGTGTTCTGT	420
TCCCTTCTCC	CGTTTCCCGG	TCGTTTCTCC	CCGTCGCTGC	GGCCCTCTCC	CCGGTCGGTC	480
GCCCCGCGGT	GCTGCCGAGC	CCCCCTTCT	GGGGGGGATG	CCCGGGCAGC	CACGCGTCCG	540
GGCGCGCACT	GTGGTTCGGG	AGCTGCTCGG	CAGGCGGCTG	AGCCAGTTGG	AGGGGCGCTCA	600
TGCCCCCGCG	GCTCTCCGCT	CGCGACGCGG	CGTGTCTTTC	GGGGGGGCTC	GTGCGTGGCG	660
GAGAGCTGCG	CAGCTTGTGC	GTCTTTCGGA	GGGAAGAGAG	CTTTTTTTTT	TTAGGGGGTC	720
GTCCCTTGCT	GTCCCGTCGG	CGGTGGATCC	GGCCT			755

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 463 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGCCGAGGTG	CGTCTGCGGG	TTGGGGCTCG	TCCGGCCCCG	TGCTCTCTCG	GGAAGGCGTT	60
TAGCCGGTAC	CGTCGCGCGC	CGAGGTGGG	CGCAGTCTCG	TGAGATAACC	CCGAGCGTGT	120
TTCTGGTTGT	TGCGCGCGCG	GGCTCCGCTG	GATGTCTTCC	CCTCCCCCTC	TCCCCGAGGC	180
CAGGTACAGC	TCCGCTCTGG	GGCTTCGTCG	GCCGCTCTCC	CCCCCTTAC	GTCCCTCGCG	240
AGCGAGCCCG	TCCGTTTCAG	CTTCTTCCG	CCTTCCCCC	ATCTTTCGCG	GCTCCGTTGG	300
CCCCGGGGTT	TTCAAGGCGC	CCCCACGCT	CCTCCGCTC	TCCGCCCCGT	GTTTGACGCG	360
CTGGTTCCCG	TCTCCCCGCG	AAACCCCGGT	TGGGTGTGTC	TCCGGCCCCG	GCTTGCTCTT	420
CGGGTCTCCC	AACCCCGCGC	CGGAAGGTTT	CGGGGTTTCC	GGG		463

- (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 378 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGATTCTTCA	GGATTGAAAC	CCAAACCGGT	TCAATTCTCT	TTCCGGCTCC	GGCCGGGGGG	60
GGCCGGCCCC	GGCGGTTTGG	TGAGTTAGAT	AACCTCGGCG	CGATCGCACG	CCCCCGGTGG	120
CGGCGACGAC	CCATTGCAAC	GTCGTCCCTA	TCAACTTTCC	ATGGTAGTCG	ATGTGCTTAC	180
CATGGTGACC	ACGGGTGACG	GGGAATCAGG	GTTCGATTCC	GGAGAGGGAG	CCTGAGAAAC	240
GGCTACCACA	TCCAAGGAAG	GCAGCAGGGG	CGCAAATTAC	CCACTCCCGA	CCCGGGGAGG	300
TAGTGACGAA	AAATAACAAT	ACAGGACTCT	TTCCAGGCCC	TGTAATTGGA	ATGAGTCCAC	360
TTAAATCTCT	TTAAGCACG					378

- (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 378 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
 (iii) HYPOTHETICAL: NO
 (iv) ANTISENSE: NO
 (v) FRAGMENT TYPE:
 (vi) ORIGINAL SOURCE:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GATCCAPTGG	AGGGCAAGTC	TGGTGCCAGC	AGCCGCGGTA	ATTCCAGCTC	CAATAGCGTA	60
TATTAAGATT	GCTCAAGTTA	AAAAGCTCGT	AGTTGGATCT	TGGGAGCGGG	CGGGCGGTCC	120
CGCGGAGGGC	GAGTCACCGC	CCGTCCCCCG	CCCTGCTCTC	TCGGCGCCCC	CTCGATGCTC	180

TTAGCTGAGT	TGTCCTCCGCG	GGCCCCGAAGC	GTTTACTTTG	AAAAAATTAG	AGTTGTTTCA	240
AAGCAGGCC	GAGCCGCTG	GATACCGCA	GCTAGGAAAT	AATGGAATAG	GACCCGCGTT	300
CCTATTTTGT	TTGGTTTTCG	GAAGTGAAGC	CATGATTAA	GGAAACGGCC	GGGGCGATT	360
CCTTATTGCG	CCCCCTA					378

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 719 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGATCTTTC	CGCTCCCGT	TCCTCCCGC	CCCTCCACCC	CGCGCTCC	CCCCCTCTT	60
TCCTCTCTC	GGAGGGGGG	GAGGTGGGG	CGCGTGGGG	GGGTCCGGG	TGGGGTCCGG	120
GGGGGACCG	CCCCGGCCG	CAAAAGCCG	CCCGCGGGG	CACCTCAAC	GTAGCGGTGC	180
GGCGGACCG	GCTACGAGC	GGCTGGGAAG	GCCCCAGCG	GAATGTGGCT	CGGGGGGGG	240
GGCGGCTCT	AGGGCGCGC	GAACCACTC	ACCCCGAGT	TTACAGCCCT	CGGGCCGGC	300
TTTCGCGAA	TCCCGGGGC	GAGGGGAAG	CCGATACCG	TCGCCCGCT	TTTCCCTCC	360
CCCGCTCCG	CTCCCGGGC	GGCGTGGGG	TGGGGGGGG	GCCGCCCTC	CCACGCCCGT	420
GGTTCTCTC	TCCTCCGGT	TCGGCCGGT	TGGGGGGGG	AGCCCGGTT	GGGGCGGGG	480
GGACTGTCT	CAGTCGCGC	CGGGCGTCT	CGCGCCGTC	GGCCCGGGG	GTCTCTCGG	540
TCACGCGCC	CCCGACGAG	CCGAGCGAC	GGGGTCGGC	CGGATGTGG	CTACCCACCC	600
GACCCGCTT	GAAACACGA	CAAAGGAGT	TAACCGGTG	CGGAGTCAG	GGCTCGCAC	660
AAGCGCTTC	TGGCGCAAT	AAGGTGAAG	GCCCCCGCG	GGGGCCCGG	GTGGGATCC	719

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 685 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CGAGGCTCT	CCAGTCCGC	GAGGGCGAC	CACCGGCCG	TCTCGCCCG	CGCGTCGGG	60
AGGTGGAGCA	CGAGCGTAC	CGTTAGAAC	CGAAAGATG	TGAACATGC	CTGGGCAGG	120
CGAAGCCAGA	GGAAGCTCT	GTTGAGTCC	GTAGCGGTCC	GTACGTGCA	ATCGGTCTG	180
CGACCTGGGT	ATAGGGCGA	AAGACTAAT	GAACATCTA	GTAGCTGGT	CCCTCCGAAG	240
TTTCCTCGAT	GATAGCTGC	GCTCTCGCA	CCTTCGGAAG	CAGTTTATC	CGGGTAAAGG	300
CGAATCGAT	TAGAGGTCT	TGGGGCGGA	AACGATCTA	AACATTTCT	CAAACTTTAA	360
ATCGGTAAAG	AAGCCCGCT	CGCTCGCGT	GAGCCGGGC	TGGAATGCA	GTGCCTAGT	420
GGCCACTTTT	GGTAAGCAG	ACTGGCGCT	CGGGATGAAC	CGAACCGCG	GTAAAGGCG	480
CCGATCGCGA	CGCTCATCAG	ACCCAGAA	AGGTGTTGT	TGATATAGC	AGCAGGACG	540
TGCCATAGGA	AGTCTGGAAT	CGCTAAGAG	TGTGTAACAA	CTACCTGCC	GAATCAACTA	600
GCCTGAA	TGATGGCGC	TGGAGCGTC	GGCCCATACC	CGGCCGTGC	CGGCAGTCG	660
AACGGGACG	GACGGGAGC	GCCG				685

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GAGGAATTCC CCTATCCCTA ATCCAGATTG GTG

33

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AAACTGCAGG CCGAGCCACC TCTCTTGT GTTG

35

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AGGAATTCAC AGAAGAGAGG TGGCTCGGCC TGC

33

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AGCCTGCAGG AAGTCATACC TGGGGAGGTG GCCC

34

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 80 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AAACTGCAGG TTAATTAACC CTAACCTAA CCCTAACCT AACCTAAC CTAACCTAA
CCCTAACCT AACCCGGAT

60
80

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTGGGCCCTA GGCTTAAGG

19

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE:

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCCAGGGTTT TCCAGTCAC GACGT

25

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCTGCAAGGC GATTAAGTTG GGTAAC

26

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TATGTTGTGT GGAATTGTGA GCGGAT

26

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: Genomic DNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTISENSE: NO
- (v) FRAGMENT TYPE:
- (vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GGGTTTAAAC AGATCTCTGC A

21

WHAT IS CLAIMED:

1. A method for producing a cell that contains heterologous nucleic acid, comprising:
 - introducing a satellite artificial chromosome (SATAC) into the cell.
- 5 2. The method of claim 1, wherein the cell is an animal cell.
3. The method of claim 2, wherein the cell is a mammalian cell.
4. The method of claim 1, wherein the SATAC is a plant artificial chromosome.
5. The method of claim 1, wherein the SATAC is a mammalian artificial chromosome.
- 10 6. The method of claim 1, wherein the cell is a plant cell.
7. The method of claim 6, wherein the SATAC is a plant artificial chromosome.
8. The method of claim 6, wherein the SATAC is a mammalian artificial chromosome.
- 15 9. The method of claim 1, wherein the SATAC is introduced by a method selected from direct DNA transfer or uptake, electroporation, lipofection, nuclear microinjection, liposome mediated transfer, microprojectile bombardment, polycation-mediated transfer and microcell fusion.
- 20 10. The method of claim 1, wherein the SATAC is introduced by a method selected from microinjection, nuclear transfer, electrofusion, projectile bombardment and lipid-mediated transfer systems.
11. The method of claim 1, wherein the SATAC is produced by a method, comprising:
 - introducing a DNA fragment into a cell, wherein the DNA fragment comprises a selectable marker;
 - growing the cell under selective conditions to produce cells that have incorporated the DNA fragment into their genomic DNA; and
 - 25 selecting a cell that comprises a satellite artificial chromosome (SATAC).

12. The method of claim 11, wherein the DNA fragment is introduced into or adjacent to an amplifiable region of a chromosome in the cell.
13. The method of claim 12, wherein the amplifiable region
5 comprises rDNA.
14. The method of claim 12, wherein the amplifiable region comprises heterochromatin.
15. The method of claim 11, wherein the DNA is introduced into pericentric heterochromatin in a chromosome of the cell.
- 10 16. The method of claim 11, wherein the cell is a mammalian cell or a plant cell.
17. The method of claim 11, wherein the method for producing the SATAC further comprises isolating the SATAC.
18. The method of claim 11, wherein in the method for
15 producing the SATAC, the DNA fragment comprises a sequence of nucleotides that targets the fragment to the heterochromatic region of a chromosome.
19. The method of claim 18, wherein the targeting sequence of nucleotides comprises satellite DNA.
- 20 20. The method of claim 12, wherein, in the method for producing the SATAC, the cell is a mammalian cell, a plant cell a fish cell, an insect cell, a reptilian cell, and amphibian cell, an arachnid cell or a rodent cell.
21. The method of claim 1, wherein the SATAC is an isolated
25 substantially pure SATAC.
22. The method of claim 1, wherein the SATAC is a megachromosome, comprising about 50 to about 450 megabases (Mb).
23. The method of claim 1, wherein the SATAC is a megachromosome, comprising about 250 to about 400 Mb.
- 30 24. The method of claim 1, wherein the SATAC is a megachromosome, comprising about 150 to about 200 Mb.

25. The method of claim 1, wherein the SATAC is a megachromosome, comprising about 90 to about 120 Mb.

26. The method of claim 1, wherein the SATAC is a megachromosome, comprising about 15 to about 60 Mb.

5 27. The method of claim 1, wherein the cell is an embryonic cell.

28. The method of claim 1, wherein the cell is an embryonic stem cell.

29. The method of claim 4, wherein the plant is tobacco, rice, maize, rye, soybean, Brassica napus, cotton, lettuce, potato, tomato or
10 arabidopsis.

30. The method of claim 6, wherein the plant is tobacco, rice, maize, rye, soybean, Brassica napus, cotton, lettuce, potato, tomato or arabidopsis.

31. The method of claim 1, wherein the SATAC comprises
15 nucleic acid encoding a gene product or a plurality of gene products.

32. The method of claim 31, wherein the gene products comprise a biosynthetic pathway.

33. The method of claim 2, wherein the animal is a mammal, bird, fowl, fish, invertebrate, vertebrate, reptile or insect.

ABSTRACT

Methods for preparing cell lines that contain artificial chromosomes, methods for preparation of artificial chromosomes, methods for purification of artificial chromosomes, methods for targeted insertion of
5 heterologous DNA into artificial chromosomes, and methods for delivery of the chromosomes to selected cells and tissues are provided. Also provided are cell lines for use in the methods, and cell lines and chromosomes produced by the methods. Methods for use of the artificial chromosomes are also provided.

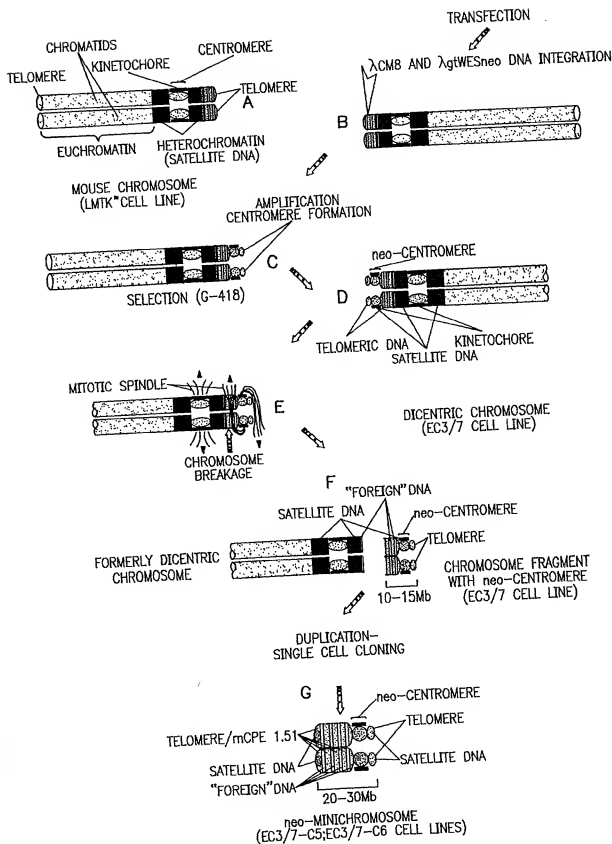


FIG. 1

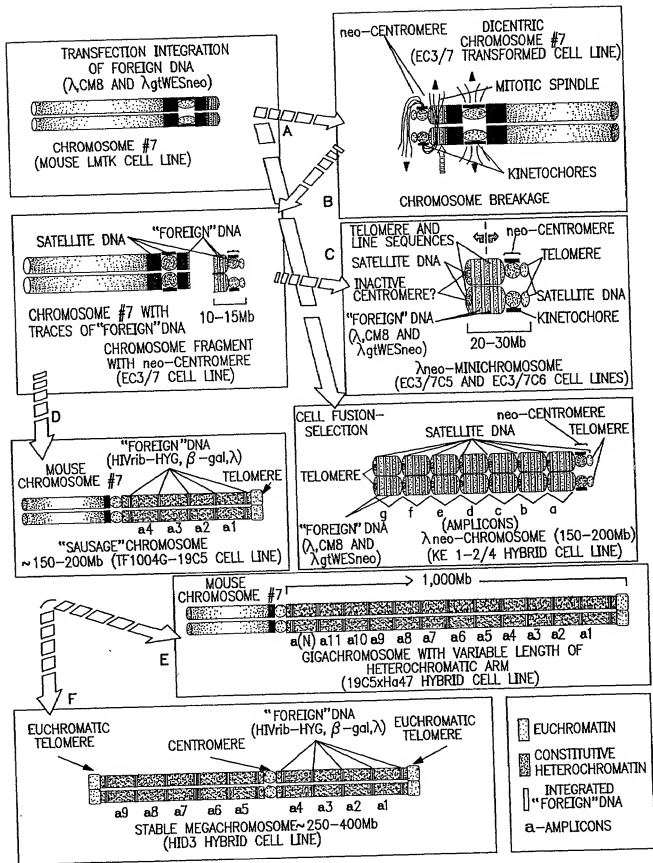


FIG. 2

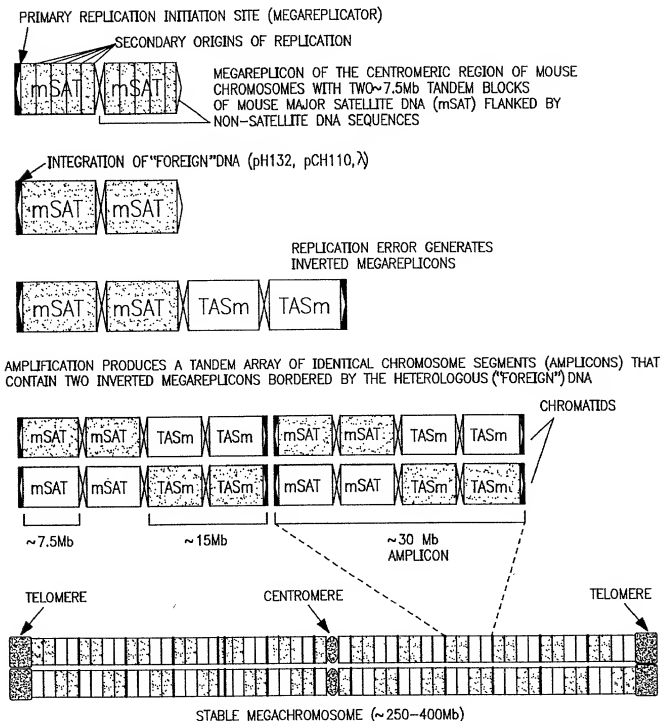


FIG. 3

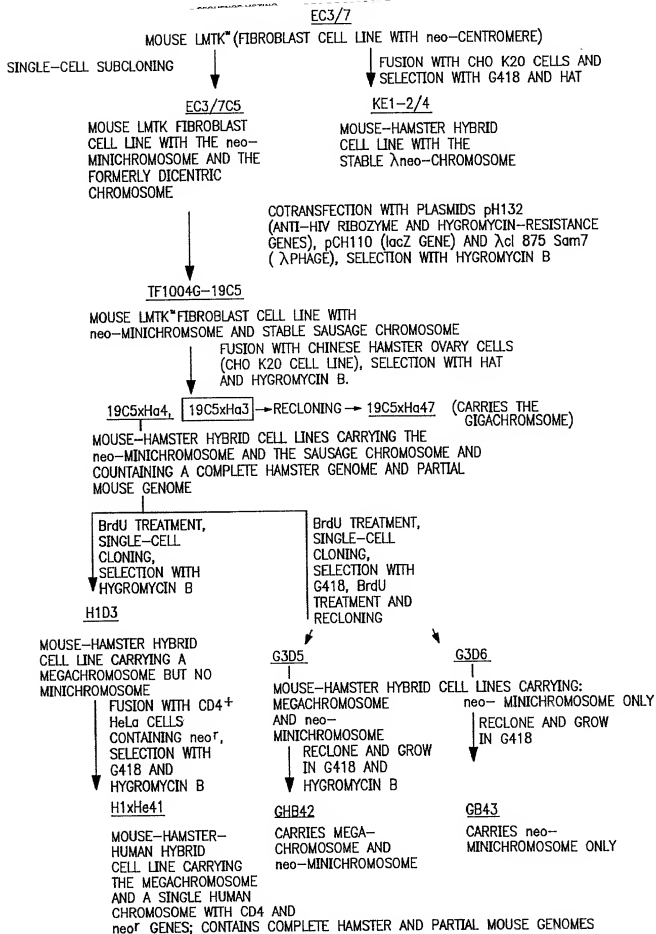


FIG. 4

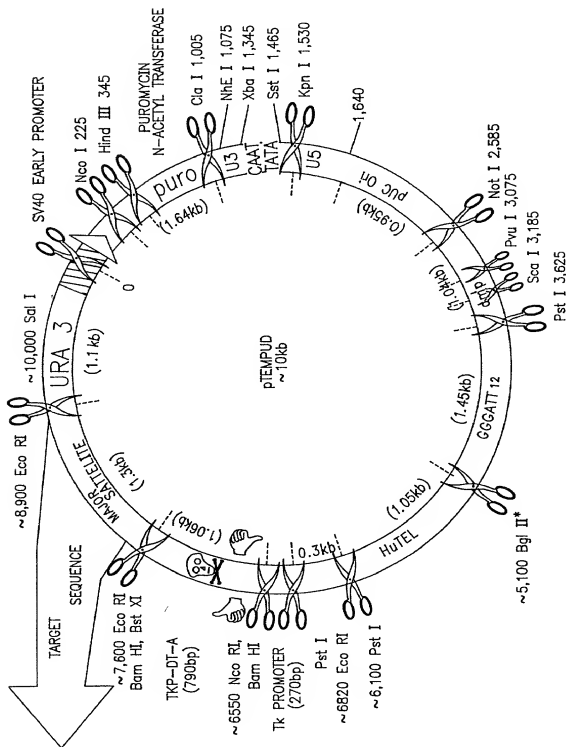


FIG. 5

DECLARATION FOR PATENT APPLICATION

As a below-named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

the specification of which (check one)

☐ is attached hereto.

☒ was filed by an authorized person on my behalf on
April 10, 1997 as Application Serial No. 08/835,682.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate listed below and so identified, or §365(a) of any PCT international application that designated at least one country other than the United States of America, listed below, and I have also identified below any foreign application for patent or inventor's certificate or PCT international application on this invention filed by us or our legal representatives or assigns and having a filing date before that of the application on which priority is claimed.

<u>Number</u>	<u>Country</u>	<u>Day/Month/Year Filed</u>	<u>Priority Claimed (Yes or No)</u>
N/A			

I hereby claim benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below:

<u>Application Serial No.</u>	<u>Filing Date</u>
N/A	

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application Serial No.

08/695,191
08/682,080
08/629,822

Filing Date

08/07/96
07/15/96
04/10/96

Status

Pending
Pending
Pending

PCT Application Serial No.

N/A

Filing Date

Status

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

I hereby appoint the following attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the United States Patent and Trademark Office connected therewith and request that all correspondence and telephone calls in respect to this application be directed to Stephanie Seidman, BROWN, MARTIN, HALLER & McCLAIN, 1660 Union Street, San Diego, California 92101-2926:

Attorney

Reg. No.

Stephanie Seidman
Carl Brown
Neil R. Martin
John L. Haller
James W. McClain
Paula K. Schoeneck

33,779
18,587
23,088
27,795
24,536
39,362

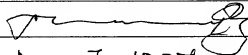
Address for correspondence:

Stephanie Seidman
BROWN, MARTIN, HALLER & McCLAIN
1660 Union Street
San Diego, CA 92101-2926

Full name joint inventor:

GYULA HADLACZKY

Inventor's signature:



Date:

May 7, 1997

Residence:

HUNGARY

Post Office Address:

H. 6723, SZEGED
SZAMOS U.1.A. IX. 36.
HUNGARY

Citizenship:

HUNGARIAN

Full name of joint inventor:

ALADAR A. SZALAY

Inventor's signature:

A. D. Galay

Date:

May 8th

Residence:

HIGHLAND, CALIFORNIA

Post Office Address:

7327 FAIRWOOD

HIGHLAND, CALIFORNIA 92346

Citizenship:

UNITED STATES OF AMERICA AND

CANADA

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: HADLACZKY et al.
 Serial No.: 08/835,682 Group Art Unit: 1804
 Filed: April 10, 1997 Examiner: Unassigned
 For: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR
 PREPARING ARTIFICIAL CHROMOSOMES

POWER OF ATTORNEY BY ASSIGNEE

Assistant Commissioner for Patents
 Washington, D.C. 20231

Sir:

Revoking any and all powers of attorney and authorizations of agent heretofore made in respect to this application, an Assignee of the above-captioned patent application hereby appoints the following as its attorneys, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

STEPHANIE SEIDMAN, Registration No. 33,779; PAULA K. SCHOENECK, Registration No. 39,362; and BARRY S. WILSON, Registration No. 39,431. Direct all telephonic inquiries to Stephanie Seidman at telephone no. (619) 450-8400.

Address all correspondence to:

Stephanie L. Seidman
 Heller Ehrman White & McAuliffe
 4250 Executive Square, 7th Floor
 La Jolla, California 92037-9103
 Telephone No. (619) 450-8400.

THE BIOLOGICAL RESEARCH CENTER OF THE
 HUNGARIAN ACADEMY OF SCIENCES

Date: 21.09.93

By: [Signature]

Name: RASKO ISTVAN M.D.
 Title: ASS. GENERAL DIRECTOR

By: [Signature]

Name: Vályi László Ph.D.
 Title: Financial Director



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: HADLACZKY et al.
Serial No.: 08/835,682 Group Art Unit: 1804
Filed: April 10, 1997 Examiner: Unassigned
For: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR
PREPARING ARTIFICIAL CHROMOSOMES

CERTIFICATE UNDER 37 C.F.R. §3.73(b)

Assistant Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

THE BIOLOGICAL RESEARCH CENTER OF THE HUNGARIAN ACADEMY OF SCIENCES, organized under the laws of Hungary with its principal place of business at H-6701 Szeged, P.O.B. 521, Hungary, is an Assignee in the above-captioned patent application by virtue of an assignment from inventor Hadlaczký. The Assignment from the inventor was recorded in the U.S. Patent and Trademark Office at Reel 8618, Frame 0716.


The undersigned (whose title is supplied below) is empowered to act on behalf of the assignee.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

THE BIOLOGICAL RESEARCH CENTER OF THE HUNGARIAN
ACADEMY OF SCIENCES

Date: 21.09.98

By:


Name: ISTVÁN BALÓ MD, PhD
Title: ASS. GENERAL DIRECTOR